

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:02 ; Search time 177 Seconds
(without alignments)
2638.788 Million cell updates/sec

Title: US-10-715-810-4

Perfect score: 6821
Sequence: 1 MPFVNQFNKYPNGVDIA.....EFIPVDGCGERPLHHHHH 1302

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6756	99.0	1296	2 AAR95010	Aar95010 C. botuli
2	6727	98.6	1295	5 AAU99339	Aau99339 Clostridi
3	4515	66.2	871	2 AAW56008	Aaw56008 Botulinum
4	4515	66.2	871	8 ADL90290	Adl90290 Clostridi
5	4512	66.1	871	8 ADL90437	Adl90437 Clostridi
6	4512	66.1	873	8 ADL90447	Adl90447 Clostridi
7	4512	66.1	873	8 ADL90445	Adl90445 Clostridi
8	4509	66.1	873	8 ADL90451	Adl90451 Clostridi
9	4507	66.1	871	2 AAW56019	Aaw56019 Recombina
10	4507	66.1	871	2 AAW56007	Aaw56007 Recombina
11	4507	66.1	871	8 ADL90435	Adl90435 Clostridi
12	4507	66.1	871	8 ADL90308	Adl90308 C. botuli
13	4507	66.1	871	8 ADL90284	Adl90284 Clostridi
14	4507	66.1	873	8 ADL90433	Adl90433 Clostridi
15	4507	66.1	873	8 ADL90431	Adl90431 Clostridi
16	4507	66.1	879	8 ADL90441	Adl90441 Clostridi
17	4507	66.1	1127	8 ADL90322	Adl90322 Clostridi
18	4507	66.1	1129	8 ADL90324	Adl90324 Clostridi
19	4504	66.0	873	8 ADL90449	Adl90449 Clostridi
20	4501	66.0	1013	2 AAW56013	Aaw56013 Recombina
21	4501	66.0	1013	8 ADL90300	Adl90300 Clostridi
22	4495.5	65.9	1130	8 ADL90326	Adl90326 Clostridi
23	4495.5	65.9	1130	8 ADL90421	Adl90421 Clostridi
24	4495.5	65.9	1132	8 ADL90328	Adl90328 Clostridi
25	4495	65.9	873	2 AAW56016	Aaw56016 Recombina

26	4495	65.9	873	8 ADL90288	Adl90288 C. botuli
27	4495	65.9	875	2 AAW56009	Aaw56009 Recombina
28	4495	65.9	875	8 ADL90292	Adl90292 Clostridi
29	4495	65.9	877	8 ADL90439	Adl90439 Clostridi
30	4495	65.9	887	8 ADL90443	Adl90443 Clostridi
31	4495	65.9	894	2 AAW56015	Aaw56015 Recombina
32	4495	65.9	894	2 ADL90286	Adl90286 Clostridi
33	4494.5	65.9	907	2 AAW56012	Aaw56012 Recombina
34	4494.5	65.9	907	8 ADL90298	Adl90298 Clostridi
35	4494.5	65.9	953	2 AAW56011	Aaw56011 Recombina
36	4494.5	65.9	953	8 ADL90296	Adl90296 Clostridi
37	4493.5	65.9	878	2 AAW56010	Aaw56010 Recombina
38	4493.5	65.9	878	8 ADL90294	Adl90294 Clostridi
39	4485	65.8	1127	8 ADL90340	Adl90340 Clostridi
40	4485	65.8	1129	8 ADL90338	Adl90338 Clostridi
41	4484.5	65.7	908	8 ADL90346	Adl90346 Clostridi
42	4479	65.7	1127	8 ADL90332	Adl90332 Clostridi
43	4479	65.7	1129	8 ADL90330	Adl90330 Clostridi
44	4473.5	65.6	949	8 ADL90350	Adl90350 Clostridi
45	4473	65.6	914	8 ADL90342	Adl90342 Clostridi

ALIGNMENTS

RESULT 1
AAR95010
ID AAR95010 standard; protein; 1296 AA.

XX AC AAR95010;
XX DT 09-JUL-1996 (first entry)
XX DE C. botulinum type A neurotoxin.
XX KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.
XX OS Clostridium botulinum.
XX PN WO9612802-A1.
XX PD 02-MAY-1996.
XX PF 23-OCT-1995; 95WO-US013737.
XX PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
XX (OPHI-) OPHIDIAN PHARM INC.

Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
Firca JR;

WPI; 1996-230603/23.
N-PSDB; AAT29244.

Fusion proteins comprising non-toxin protein and part of toxin - useful to form anti-toxins against Clostridium botulinum type A, and C. difficile type toxins, and to treat C. difficile intoxication, partic. diarrhoea.

Claim 4; Page 344-350; 434pp; English.

Clostridium botulinum type A neurotoxin (AAR95010) is processed to form a dimer composed of a light and a heavy chain. It is the product of the type A neurotoxin gene (AAT29244). The 50 kDa C-terminal portion of the heavy chain, or C fragment (see also AAR95008), was produced using a synthetic gene (AAT29245) having codon usage altered to improve expression in Escherichia coli. Fusion proteins of the type A toxin or C fragment, with e.g. maltose binding protein or polyhistidine affinity tag (see also AAR95008), are used to generate neutralising antitoxins and in

XX 27-NOV-2001; 2001WO-US045059.
 XX PF
 XX PR
 XX PA (ALLR) ALLERGAN SALES INC.
 XX PI Lin W, Aoki KR, Steward LE;
 XX WPI; 2002-557531/59.
 XX Novel modified botulinum toxin or tetanus toxin comprising a protease
 PT cleavage site, is useful for treating conditions benefited by neurotoxin
 PT activity.
 XX
 XX Disclosure; Fig 1; 19pp; English.
 XX
 CC The invention discloses modified botulinum toxin (BoToX) or tetanus toxin
 CC (TeToX) which contain new protease cleavage sites. These sites are
 CC created in regions where the cleavage site is specifically susceptible to
 CC cleavage by a protease that is present in effective levels only in a
 CC tissue where toxicity is undesirable and where the cleavage site itself
 CC does not inactivate the toxin. BoToX and TeToX are neurotoxins which are
 CC highly potent and specific poisons of neural cells. The toxins are
 CC synthesised in vivo as single chains, which are not toxic, but become
 CC active when nicked, in a post-translational modification, to form
 CC separate light and heavy chains which are linked through a disulphide
 CC bond. The modified toxins are useful to treat conditions benefited by
 CC neurotoxin activity (e.g. spastic conditions, including stabismus,
 CC bophliarospasm and hemifacial spasm, brain injury, spinal cord injury,
 CC stroke, multiple sclerosis and cerebral palsy) by administering the
 CC toxins for their localised production. The advantage of the toxins are
 CC that they are deactivated in tissues where toxic activity is undesirable
 CC and activated at desired targets. The sequence presented is the
 CC Clostridium botulinum toxin A toxin (BoNT/A) protein
 XX
 SQ Sequence 1295 AA;
 Query Match 98.6%; Score 6727; DB 5; Length 1295;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1288; Conservative 1; Mismatches 6; Indels. 0; Gaps 0;
 QY 2 PFVNKQPNYKDPVNGVDIAIKIPNAGOMQPVKAFKHNIKIWIIPERDFTTNBEGDLNP 61
 DB 1 PFVNKQPNYKDPVNGVDIAIKIPNAGOMQPVKAFKHNIKIWIIPERDFTTNBEGDLNP 60
 QY 62 PPEAKQPVSVSYDSTYLTSTNEKDNLYLKVYTKLPERIYSTDLGRMLTSTIVRGIPFWGGS 121
 DB 61 PPEAKQPVSVSYDSTYLTSTNEKDNLYLKVYTKLPERIYSTDLGRMLTSTIVRGIPFWGGS 120
 QY 122 TIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSPGHEVLNLRNGYG 181
 DB 121 TIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSPGHEVLNLRNGYG 180
 QY 182 STQYIRFSPDFTFGFEESLEVDTNPLLCAGKFAFDPAVTLAHELIIYAGHRLYGIAPNPR 241
 DB 181 STQYIRFSPDFTFGFEESLEVDTNPLLCAGKFAFDPAVTLAHELIIYAGHRLYGIAPNPR 240
 QY 242 VFKNTHAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKPDIASITLNKAK 301
 DB 241 VFKNTHAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKPDIASITLNKAK 300
 QY 302 SIVGTASLQVMKNVFEKYLSDTSKFSVDKLKFDKLYKMLTEIYTEDNFVKPFKVL 361
 DB 301 SIVGTASLQVMKNVFEKYLSDTSKFSVDKLKFDKLYKMLTEIYTEDNFVKPFKVL 360
 QY 362 NRKTYLNFDAKFAVKINIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFTG 421
 DB 361 NRKTYLNFDAKFAVKINIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFTG 420
 QY 422 LFEFYKLLCVRGIITSTKTSKLDKGNKALNDLCIKVNNWDLFFSPSEDFTNDLNKGEI 481
 DB 421 LFEFYKLLCVRGIITSTKTSKLDKGNKALNDLCIKVNNWDLFFSPSEDFTNDLNKGEI 480

482 TSDTNEAAEENISLDLIQOYLYTFNFDNEPENISLENSSDIIGOLELMPNIEFPNGK 541
 481 TSDTNEAAEENISLDLIQOYLYTFNFDNEPENISLENSSDIIGOLELMPNIEFPNGK 540
 542 KYELDKYTFPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKYNKATEAA 601
 541 KYELDKYTFPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKYNKATEAA 600
 602 MFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAIINMGLYKODFVGALIFSGA 661
 601 MFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPAIINMGLYKODFVGALIFSGA 660
 662 VILLEPIETAIPVLGTFALVSYIANKVLTVQTDNALSKRNEKWEVYKIIVTNMLAKV 721
 661 VILLEPIETAIPVLGTFALVSYIANKVLTVQTDNALSKRNEKWEVYKIIVTNMLAKV 720
 722 NTQIDILIRKMKALENQAEATKAIINYQYNOVTESEKNNINENIDLSKLNESINKAM 781
 721 NTQIDILIRKMKALENQAEATKAIINYQYNOVTESEKNNINENIDLSKLNESINKAM 780
 782 ININKELNOCVSYLNMNSMIPYGVKLEDFDASLKDALLKVIYDNRGTLIGQVDRDKDV 841
 781 ININKELNOCVSYLNMNSMIPYGVKLEDFDASLKDALLKVIYDNRGTLIGQVDRDKDV 840
 842 NNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 901
 841 NNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 900
 902 SKYVFPDIDKNOQLNLSSEKIEVLKNAIVNMYENFSTFWIRIPKYFNSISLANE 961
 901 SKYVFPDIDKNOQLNLSSEKIEVLKNAIVNMYENFSTFWIRIPKYFNSISLANE 960
 962 YTIINCWENNSGKWSLVNIGEIIWTLDQTOEIKORVVFKYSOMINISDIYINRIFVTITN 1021
 961 YTIINCWENNSGKWSLVNIGEIIWTLDQTOEIKORVVFKYSOMINISDIYINRIFVTITN 1020
 1022 NRLNNSKIYINGRLIDQKPIISNLGNITHASNNIMPKLDGCRDTHRYIWKYFNLFDKELNE 1081
 1021 NRLNNSKIYINGRLIDQKPIISNLGNITHASNNIMPKLDGCRDTHRYIWKYFNLFDKELNE 1080
 1082 KEIKDLIDNQSNGIILKDFWGDYLDVDPKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRG 1141
 1081 KEIKDLIDNQSNGIILKDFWGDYLDVDPKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRG 1140
 1142 SVMTTNLYNLSLYRGTKFTIKKYASGNKDNIVRNNDRVYINVVVKNYRRLATNASQAG 1201
 1141 SVMTTNLYNLSLYRGTKFTIKKYASGNKDNIVRNNDRVYINVVVKNYRRLATNASQV 1200
 1202 VEKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNGNDIGFIGHQFNNTAKL 1261
 1201 VFKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNGNDIGFIGHQFNNTAKL 1260
 1262 VASNWNRQIERSRSLGCSWEEFIPVDDGWERPL 1296
 1261 VASNWNRQIERSRSLGCSWEEFIPVDDGWERPL 1295
 RESULT 3
 AAW56008
 ID AAW56008 standard; protein; 871 AA.
 XX AC AAW56008;
 XX AC
 XX DT 27-JUL-1998 (first entry)
 XX DE Botulinum neurotoxin type A BoNT/A.
 XX KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
 XX KW detection; tetanus; non-toxic; toxin.
 XX OS Clostridium botulinum.
 XX OS

PN W09807864-A1.
 XX
 PD 26-FEB-1998.
 XX
 PF 22-AUG-1997; 97WO-GB002273.
 XX
 PR 23-AUG-1996; 96GB-00017671.
 PR 13-DEC-1996; 96GB-00025996.
 XX
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PA (SPEY-) SPEYWOOD LAB LTD.
 XX
 PI Shone CC, Quinn CP, Foster KA;
 XX
 XX WPI: 1998-169168/15.
 DR N-PSDB; AAV26280.
 DR
 XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
 PT immunogens or as non-toxic standards for the detection of neurotoxins.
 XX
 PS Disclosure; Page 52-54; 137pp; English.
 XX
 CC The present sequence represents botulinum neurotoxin type A from the
 CC present invention. The present invention describes recombinant neurotoxin
 CC proteins which comprise a first and second domain, where the first domain
 CC is adapted to cleave one or more vesicle or plasma-membrane associated
 CC proteins essential to exocytosis, and where the second domain is adapted:
 CC (a) to translocate the protein into a cell; (b) to increase the
 CC solubility of the protein compared to the solubility of the first domain
 CC on its own, or (c) both to translocate the protein into a cell and to
 CC increase the solubility of the protein compared to the solubility of the
 CC first domain on its own, the protein being free of clostridial neurotoxin
 CC (CN) and free of CN precursor that can be converted into toxin by
 CC proteolytic action. The recombinant proteins can be used as therapeutic
 CC agents for targeting cells expressing a relevant substrate. The products
 CC can also be used as immunogens and as non-toxic standards for the
 CC assessment and development of in vitro assays for the detection of
 CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
 CC environmental samples
 XX
 SQ Sequence 871 AA;
 Query Match 66.2%; Score 4515; DB 2; Length 871;
 Best Local Similarity 99.9%; Pred. No. 4e-257;
 Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPFVNKQNPYDVPNGVDIAVIKIPNAGOMQPVKAFKIHNKIWIPIERDFTNPEEGDLN 60
 DB 1 MPFVNKQNPYDVPNGVDIAVIKIPNAGOMQPVKAFKIHNKIWIPIERDFTNPEEGDLN 60
 QY 61 PPEAKQVPVSYDYSTYLSSTNEXDNVYKVTGLFERIYSTDLGRMLITSIVRGIPFWGG 120
 DB 61 PPEAKQVPVSYDYSTYLSSTNEXDNVYKVTGLFERIYSTDLGRMLITSIVRGIPFWGG 120
 QY 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIQFECKSPGHEVLNLTNGY 180
 DB 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIQFECKSPGHEVLNLTNGY 180
 QY 181 GSTQYIRSPDFTGFESELSVDNTPGLGACKFATDPAVTLAHLIVAGHELVIAGINPN 240
 DB 181 GSTQYIRSPDFTGFESELSVDNTPGLGACKFATDPAVTLAHLIVAGHELVIAGINPN 240
 QY 241 RVFKNTNAYYEMSGLEVSFELRTFGGHDAKFIDSLQENEFRLYYNKFMDIASTLNKA 300
 DB 241 RVFKNTNAYYEMSGLEVSFELRTFGGHDAKFIDSLQENEFRLYYNKFMDIASTLNKA 300
 QY 301 KSIYGTASLQYKMNKVFKEKYLLEDSTSGKFSVDKLPDKLYKMLTEIYTDNDFVKFKV 360
 DB 301 KSIYGTASLQYKMNKVFKEKYLLEDSTSGKFSVDKLPDKLYKMLTEIYTDNDFVKFKV 360
 QY 361 LNRKTYLNFDKAVKINIVPKVNTIYDGNLRNTNLAANFGONTNINNMFKLKNFT 420
 DB 361 LNRKTYLNFDKAVKINIVPKVNTIYDGNLRNTNLAANFGONTNINNMFKLKNFT 420

421 GLFEFYKLLCVRGIIITSKTSKSLDKGVNKAINDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
 421 GLFEFYKLLCVRGIIITSKTSKSLDKGVNKAINDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
 481 IISDTNIEAAENISLDLIQOYVLTFFNENPENISNIENISLDIIIGOLELMPNIEFPNG 540
 481 IISDTNIEAAENISLDLIQOYVLTFFNENPENISNIENISLDIIIGOLELMPNIEFPNG 540
 541 KKYELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKKVNKATEA 600
 541 KKYELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKKVNKATEA 600
 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGLIFSG 660
 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGLIFSG 660
 661 AVTLLEFIPEIATPVLTGTFALSVYIANKVLTWTIDNALSKRNEKWDVVKYIVTNWLAK 720
 661 AVTLLEFIPEIATPVLTGTFALSVYIANKVLTWTIDNALSKRNEKWDVVKYIVTNWLAK 720
 721 VNTQIDILIRKMKKEALENQAATKAIINYQNOYTEEKNINFNIDDLSSKLNESINKA 780
 721 VNTQIDILIRKMKKEALENQAATKAIINYQNOYTEEKNINFNIDDLSSKLNESINKA 780
 781 MININKPLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 840
 781 MININKPLNOCSSVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 840
 841 VNNLTSTDIPFQLSKYVDNORLLSTFTYIK 871
 841 VNNLTSTDIPFQLSKYVDNORLLSTFTYIK 871

RESULT 4
 ADL90290
 ID ADL90290 standard; protein; 871 AA.
 XX
 AC ADL90290;
 XX
 DT 17-JUN-2004 (first entry)
 XX
 DE Clostridium botulinum BoNT/A protein SEQ ID NO:8.
 XX
 KW single chain polypeptide; clostridial neurotoxin light chain;
 KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
 KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
 KW botulinum; tetanus.
 XX
 OS Clostridium botulinum.
 XX
 PN W02004024909-A2.
 XX
 PD 25-MAR-2004.
 XX
 PF 12-SEP-2003; 2003WO-GB003824.
 XX
 PR 12-SEP-2002; 2002US-00241596.
 XX
 XX (HEAL-) HEALTH PROTECTION AGENCY.
 XX
 PI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 PI Wayne J;
 XX
 XX WPI: 2004-270039/25.
 DR N-PSDB; ADL90289.
 DR
 XX New single chain polypeptides comprising clostridial neurotoxin light and
 PT heavy chains, useful as positive controls for toxin assays, or for
 PT developing vaccines against clostridial toxin.
 XX
 PS Disclosure; SEQ ID NO 8; 588pp; English.
 XX

CC The present invention describes a single chain polypeptide comprising
CC clostridial neurotoxin light and heavy chains. The single chain
CC polypeptide comprises 2 domains: the first domain is a clostridial
CC neurotoxin light chain, or its fragment or variant, which is capable of
CC cleaving one or more vesicle or plasma membrane associated proteins
CC essential to exocytosis; the second domain is a clostridial neurotoxin
CC heavy chain H-N portion, or its fragment or variant, which is capable of
CC translocating the polypeptide into a cell and/or increasing the
CC solubility of the polypeptide compared to the solubility of the first
CC domain on its own. The second domain lacks a functional C-terminal part
CC of a clostridial neurotoxin heavy chain, designated H-C, which renders
CC the polypeptide incapable of binding to cell surface receptors that are
CC the natural cell surface receptors to which native clostridial neurotoxin
CC binds. Also described is a nucleic acid molecule encoding the single
CC chain polypeptide described above. The single chain polypeptide has
CC antibacterial activity, and can be used in vaccines. The single chain
CC polypeptides can be used as positive controls for toxin assays, as
CC reagent components for the synthesis of therapeutic molecules, or for
CC developing vaccines against clostridial toxin. The polypeptides are also
CC useful as non-toxic standards for the assessment and development of in
CC vitro assays for detection of functional botulinum or tetanus neurotoxins
CC in foodstuffs or environmental samples. The present sequence is used in
CC the exemplification of the present invention.

XX SQ Sequence 871 AA;

Query Match 66.2%; Score 4515; DB 8; Length 871;
Best Local Similarity 99.9%; Pred. No. 4e-257;
Matches 870; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKDPVNGVDIAIKIPNAGQMPQVKAFLHKKIWIPIRDTFTNPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAIKIPNAGQMPQVKAFLHKKIWIPIRDTFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTFLERYSTDGLRMLTSLVRGIPFWGG 120
Db 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTFLERYSTDGLRMLTSLVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQFECKSGFHEVLNLTNGY 180
Qy 181 GSTOYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 181 GSTOYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYYEMSGLEVSFEELTFCGHDAKFDLSIQENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYYEMSGLEVSFEELTFCGHDAKFDLSIQENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOYKMNKVFKEKYLLEDSTSGKFSVDKLFKDLKLYKMLTEIYTDNPFVKFKV 360
Db 301 KSIIVGTTASLOYKMNKVFKEKYLLEDSTSGKFSVDKLFKDLKLYKMLTEIYTDNPFVKFKV 360
Qy 361 LNRKTYLNFDFKAVKFINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDFKAVKFINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLFFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPEDNFTNDLNKEE 480
Db 421 GLFFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPEDNFTNDLNKEE 480
Qy 481 ITSNTNTEAENISLDLIQOYIYITFTNPDNENISNENISSDIIQOELMPNTERPFG 540
Db 481 ITSNTNTEAENISLDLIQOYIYITFTNPDNENISNENISSDIIQOELMPNTERPFG 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy 601 AMFLGWELQVYDFTDETSEVSTTDKADITIIPIYGPALNIGNMLYKODFVGLIFSG 660
Db 601 AMFLGWELQVYDFTDETSEVSTTDKADITIIPIYGPALNIGNMLYKODFVGLIFSG 660

Qy 661 AVILLFPIPIAIPVLGTFAVSYIANKVLTQVOTIDNALSKRNEKWDVTKYIVTNWLAK 720
Db 661 AVILLFPIPIAIPVLGTFAVSYIANKVLTQVOTIDNALSKRNEKWDVTKYIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOYVTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOYVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 840
Db 781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 840
Qy 841 VNNLTSTIDIPQLSKYVDNQRLLSTFTTEYIK 871
Db 841 VNNLTSTIDIPQLSKYVDNQRLLSTFTTEYIK 871

RESULT 5

ADL90437
ID ADL90437 standard; protein; 871 AA.

XX AC ADL90437;

DT 17-JUN-2004 (first entry)

XX Clostridial neurotoxin amino acid sequence SEQ ID NO:155.

XX single chain polypeptide; clostridial neurotoxin light chain;
XX clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
XX antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX botulinum; tetanus.

OS Clostridium botulinum.

XX WO2004024909-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-GB003824.

XX 12-SEP-2002; 2002US-00241596.

XX (HEAL-) HEALTH PROTECTION AGENCY.

XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
XX Wayne J;

XX WPI; 2004-270039/25.

XX N-PSDB; ADL90436.

XX New single chain polypeptides comprising clostridial neurotoxin light and
XX heavy chains, useful as positive controls for toxin assays, or for
XX developing vaccines against clostridial toxin.

XX Claim 1; SEQ ID NO 155; 588pp; English.

XX The present invention describes a single chain polypeptide comprising
XX clostridial neurotoxin light and heavy chains. The single chain
XX polypeptide comprises 2 domains: the first domain is a clostridial
XX neurotoxin light chain, or its fragment or variant, which is capable of
XX cleaving one or more vesicle or plasma membrane associated proteins
XX essential to exocytosis; the second domain is a clostridial neurotoxin
XX heavy chain H-N portion, or its fragment or variant, which is capable of
XX translocating the polypeptide into a cell and/or increasing the
XX solubility of the polypeptide compared to the solubility of the first
XX domain on its own. The second domain lacks a functional C-terminal part
XX of a clostridial neurotoxin heavy chain, designated H-C, which renders
XX the polypeptide incapable of binding to cell surface receptors that are
XX the natural cell surface receptors to which native clostridial neurotoxin
XX binds. Also described is a nucleic acid molecule encoding the single
XX chain polypeptide described above. The single chain polypeptide has
XX antibacterial activity, and can be used in vaccines. The single chain

CC polypeptides can be used as positive controls for toxin assays, as
CC reagent components for the synthesis of therapeutic molecules, or for
CC developing vaccines against clostridial toxin. The polypeptides are also
CC useful as non-toxic standards for the assessment and development of in
CC vitro assays for detection of functional botulinum or tetanus neurotoxins
CC in foodstuffs or environmental samples. The present sequence is used in
CC the exemplification of the present invention.

XX SQ Sequence 871 AA;

Query Match 66.1%; Score 4512; DB 8; Length 871;
Best Local Similarity 99.9%; Pred. No. 6.le-257;
Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKHKNKIWIPIERDFTNPEEGDLN 60
Db 1 MEFVKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKHKNKIWIPIERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDSTYSLSTDNKDNLYKVTCLFERYISTDLGRMLLTSTVIRGIPFWGG 120
Db 61 PPEAKQVPVSYDSTYSLSTDNKDNLYKVTCLFERYISTDLGRMLLTSTVIRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIOPGSRSEELNLIIGPSADIIIOFECKSGFHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIOPGSRSEELNLIIGPSADIIIOFECKSGFHEVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPN 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPN 240
Qy 241 RVFKVNTNAYYEMSGLEVSFEELTFGCHDAKIDSLOENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYYEMSGLEVSFEELTFGCHDAKIDSLOENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIYVGTASLQYKQVFKVKEKLLSEDTSGKSVKDLKFDKLYKMLTIYTEDNFKVFPKV 360
Db 301 KSIYVGTASLQYKQVFKVKEKLLSEDTSGKSVKDLKFDKLYKMLTIYTEDNFKVFPKV 360
Qy 361 LNRKTYLNFDAKFKINIPKVNVTIYDGFNLRNTNLAANFNGQNTENNMFYKLNFT 420
Db 361 LNRKTYLNFDAKFKINIPKVNVTIYDGFNLRNTNLAANFNGQNTENNMFYKLNFT 420
Qy 421 GLFEPYKLLVGRGIITSTKSLDKYKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Db 421 GLFEPYKLLVGRGIITSTKSLDKYKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Qy 481 ITSDTNEAEBENISLDLIQYLYTFNFDNPEPENISLENSSDIIGOLELMPNTERFPNG 540
Db 481 ITSDTNEAEBENISLDLIQYLYTFNFDNPEPENISLENSSDIIGOLELMPNTERFPNG 540
Qy 541 KYELDXYTFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYVFFSSDYVKKYKATEA 600
Db 541 KYELDXYTFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYVFFSSDYVKKYKATEA 600
Qy 601 AMFLGWEOQLVYDFTDETSEVSTTDKIADITIIPIYGPAALNIMLYKDDFVGALLIFSG 660
Db 601 AMFLGWEOQLVYDFTDETSEVSTTDKIADITIIPIYGPAALNIMLYKDDFVGALLIFSG 660
Qy 661 AVILLEFPEIPIVGLTFALVSVYANKVLTVOIDNALSKNEKWDVEYKIVYTNMLAK 720
Db 661 AVILLEFPEIPIVGLTFALVSVYANKVLTVOIDNALSKNEKWDVEYKIVYTNMLAK 720
Qy 721 VNTQIDLRKWKKEALENOAETKAIINQYNQVTEEEKNNINFINIDLSKLNESINKA 780
Db 721 VNTQIDLRKWKKEALENOAETKAIINQYNQVTEEEKNNINFINIDLSKLNESINKA 780
Qy 781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNNTLSTDIIPQLSKYVDNORLLSTFTBYIK 871
Db 841 VNNTLSTDIIPQLSKYVDNORLLSTFTBYIK 871

RESULT 6

ADL90447

ID ADL90447 standard; protein; 873 AA.

XX AC ADL90447;

XX DT 17-JUN-2004 (first entry)

XX DE Clostridial neurotoxin amino acid sequence SEQ ID NO:165.

XX KW single chain polypeptide; clostridial neurotoxin light chain;
KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.

OS Clostridium botulinum.

XX WO2004024909-A2.

XX PD 25-MAR-2004.

XX PF 12-SEP-2003; 2003WO-GB003824.

XX PR 12-SEP-2002; 2002US-00241596.

XX PA (HEAL-) HEALTH PROTECTION AGENCY.

XX FI Stone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI Wayne J;

XX WPI; 2004-270039/25.

XX DR N-PSDB; ADL90446.

XX CC New single chain polypeptides comprising clostridial neurotoxin light and
PT heavy chains, useful as positive controls for toxin assays, or for
PT developing vaccines against clostridial toxin.

XX PS Claim 1; SEQ ID NO 165; 588pp; English.

XX CC The present invention describes a single chain polypeptide comprising
CC clostridial neurotoxin light and heavy chains. The single chain
CC polypeptide comprises 2 domains: the first domain is a clostridial
CC neurotoxin light chain, or its fragment or variant, which is capable of
CC cleaving one or more vesicle or plasma membrane associated proteins
CC essential to exocytosis; the second domain is a clostridial neurotoxin
CC heavy chain H-N portion, or its fragment or variant, which is capable of
CC translocating the polypeptide into a cell and/or increasing the first
CC solubility of the polypeptide compared to the solubility of the first
CC domain on its own. The second domain lacks a functional C-terminal part
CC of a clostridial neurotoxin heavy chain, designated H-C, which renders
CC the polypeptide incapable of binding to cell surface receptors that are
CC the natural cell surface receptors to which native clostridial neurotoxin
CC binds. Also described is a nucleic acid molecule encoding the single
CC chain polypeptide described above. The single chain polypeptide has
CC antibacterial activity, and can be used in vaccines. The single chain
CC polypeptides can be used as positive controls for toxin assays, as
CC reagent components for the synthesis of therapeutic molecules, or for
CC developing vaccines against clostridial toxin. The polypeptides are also
CC useful as non-toxic standards for the assessment and development of in
CC vitro assays for detection of functional botulinum or tetanus neurotoxins
CC in foodstuffs or environmental samples. The present sequence is used in
CC the exemplification of the present invention.

XX SQ Sequence 873 AA;

Query Match 66.1%; Score 4512; DB 8; Length 873;

Best Local Similarity 99.9%; Pred. No. 6.le-257;

Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKHKNKIWIPIERDFTNPEEGDLN 60

Db 3 MEFVNKQFNKDPVNGVDIAIYKIPNAGQMPVKAFKIHKNKIWIPIPERDTFTNPEEGDLN 62
Qy 61 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 63 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQDGSYSRSELNLIIGPSADIIQFCKSGFHEVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQDGSYSRSELNLIIGPSADIIQFCKSGFHEVLNLTNGY 182
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGHDADFIDSLQENFRLYYYNKKFKDIASLTNKA 300
Db 243 RVFKVNTNAYEMSGLEVSFEELRTFGHDADFIDSLQENFRLYYYNKKFKDIASLTNKA 302
Qy 301 KSIYVGTSTASLOYKMNVPKFKYLLSEDTSKFSVDKLFKDKLYKMLTEIYTEDNFVKPKV 360
Db 303 KSIYVGTSTASLOYKMNVPKFKYLLSEDTSKFSVDKLFKDKLYKMLTEIYTEDNFVKPKV 362
Qy 361 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNNNFTKLKNT 420
Db 363 LNRKTYLNFDFKAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNNNFTKLKNT 422
Qy 421 GLFEFYKLLCVRGITSTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 480
Db 423 GLFEFYKLLCVRGITSTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE 482
Qy 481 ITSNTIEAABENISLDLIQOYLYTFNFDNEPENISLENLSSDIIGOLELMPNTERFPNG 540
Db 483 ITSNTIEAABENISLDLIQOYLYTFNFDNEPENISLENLSSDIIGOLELMPNTERFPNG 542
Qy 541 KKYELDKYTMFHYLRQAQFEHKGKRIALTNSVNEALLNPSRVYTFSSDYVKKYKNAKTEA 600
Db 543 KKYELDKYTMFHYLRQAQFEHKGKRIALTNSVNEALLNPSRVYTFSSDYVKKYKNAKTEA 602
Qy 601 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDFVGLALIFSG 660
Db 603 AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDFVGLALIFSG 662
Qy 661 AVILLEPTEIPALVGLTFALVSIVANKVLTQVIDNALSKEKWEVYKIVYTNWLAK 720
Db 663 AVILLEPTEIPALVGLTFALVSIVANKVLTQVIDNALSKEKWEVYKIVYTNWLAK 722
Qy 721 VNTQIDLRKKWKEALENOAEATKAIINYQNVYTEBEKNNINFNIDDLSSKLNESINKA 780
Db 723 VNTQIDLRKKWKEALENOAEATKAIINYQNVYTEBEKNNINFNIDDLSSKLNESINKA 782
Qy 781 MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALKYIYDNRGTGLIGQVDRLKDK 840
Db 783 MININKFLNQCYSVLYMNSMIPYGVKRLDFDASLKDALKYIYDNRGTGLIGQVDRLKDK 842
Qy 841 VNNLTSTDIPQLSKYVDNORLLSTFTFYIK 871
Db 843 VNNLTSTDIPQLSKYVDNORLLSTFTFYIK 873

RESULT 7
ADL90445
ID ADL90445 standard; protein; 873 AA.
XX AC ADL90445;
XX DT 17-JUN-2004 (first entry)

XX DE Clostridial neurotoxin amino acid sequence SEQ ID NO:163.

XX KW single chain polypeptide; clostridial neurotoxin light chain;
XX KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.

XX OS Clostridium botulinum.
XX FN WO2004024909-A2.
XX PD 25-MAR-2004.
XX XX
XX 12-SEP-2003; 2003WO-GB003824.
XX PF 12-SEP-2002; 2002US-00241596.
XX PR (HEAL-) HEALTH PROTECTION AGENCY.
XX XX
XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
XX FI Wayne J;
XX PT WPI; 2004-270039/25.
XX DR N-PSDB; ADL90444.
XX XX
XX New single chain polypeptides comprising clostridial neurotoxin light and
XX heavy chains, useful as positive controls for toxin assays, or for
XX developing vaccines against clostridial toxin.
XX PS Claim 1; SEQ ID NO 163; 588pp; English.
XX XX
XX The present invention describes a single chain polypeptide comprising
XX clostridial neurotoxin light and heavy chains. The single chain
XX polypeptide comprises 2 domains: the first domain is a clostridial
XX neurotoxin light chain, or its fragment or variant, which is capable of
XX cleaving one or more vesicle or plasma membrane associated proteins
XX essential to exocytosis; the second domain is a clostridial neurotoxin
XX heavy chain H-N portion, or its fragment or variant, which is capable of
XX translocating the polypeptide into a cell and/or increasing the
XX solubility of the polypeptide compared to the solubility of the first
XX domain on its own. The second domain lacks a functional C-terminal part
XX of a clostridial neurotoxin heavy chain, designated H-C, which renders
XX the polypeptide incapable of binding to cell surface receptors that are
XX the natural cell surface receptors to which native clostridial neurotoxin
XX binds. Also described is a nucleic acid molecule encoding the single
XX chain polypeptide described above. The single chain polypeptide has
XX antibacterial activity, and can be used in vaccines. The single chain
XX polypeptides can be used as positive controls for toxin assays, as
XX reagent components for the synthesis of therapeutic molecules, or for
XX developing vaccines against clostridial toxin. The polypeptides are also
XX useful as non-toxic standards for the assessment and development of in
XX vitro assays for detection of functional botulinum or tetanus neurotoxins
XX in foodstuffs or environmental samples. The present sequence is used in
XX the exemplification of the present invention.
XX SQ Sequence 873 AA;

Query Match 66.1%; Score 4512; DB 8; Length 873;
Best Local Similarity 99.9%; Pred. No. 6.1e-257;
Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MPFVNKQFNKDPVNGVDIAIYKIPNAGQMPVKAFKIHKNKIWIPIPERDTFTNPEEGDLN 60
Db 3 MEFVNKQFNKDPVNGVDIAIYKIPNAGQMPVKAFKIHKNKIWIPIPERDTFTNPEEGDLN 62
Qy 61 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 63 PPPEAKQVPVSYDSTYLTSTNEKDNKYLKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQDGSYSRSELNLIIGPSADIIQFCKSGFHEVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQDGSYSRSELNLIIGPSADIIQFCKSGFHEVLNLTNGY 182
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGHDADFIDSLQENFRLYYYNKKFKDIASLTNKA 300

Db 243 RVFKVNTNAYEMSGLEVSFEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 302
Qy 301 KSIIVGTTASLQYMKNVKPEKYLLEDTSKGFSVDKLFEDKLYKMLTEIYTDNEDNFKPKV 360
Db 303 KSIIVGTTASLQYMKNVKPEKYLLEDTSKGFSVDKLFEDKLYKMLTEIYTDNEDNFKPKV 362
Qy 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQTEINNMFNFKLNFT 420
Db 363 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQTEINNMFNFKLNFT 422
Qy 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480
Db 423 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 482
Qy 481 ITSDNTIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNIERFNG 540
Db 483 ITSDNTIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNIERFNG 542
Qy 541 KKYELDKYTMPEHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Db 543 KKYELDKYTMPEHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 602
Qy 601 AMFLGWVEQLVYDFTDSEVSTTDKIADITIIPIYIGPALNIGMLYKDDFVGALLIFSG 660
Db 603 AMFLGWVEQLVYDFTDSEVSTTDKIADITIIPIYIGPALNIGMLYKDDFVGALLIFSG 662
Qy 661 AVILLEPTEIPAIVLGTFAVSVIANKVLTVQTDNALSKRNEKWDVVKYIVTNWLAK 720
Db 663 AVILLEPTEIPAIVLGTFAVSVIANKVLTVQTDNALSKRNEKWDVVKYIVTNWLAK 722
Qy 721 VNTQIDILIRKCKEALENQAEATKAIINQYQNTYEEKNNINFNIDDLSSKLNESINKA 780
Db 723 VNTQIDILIRKCKEALENQAEATKAIINQYQNTYEEKNNINFNIDDLSSKLNESINKA 782
Qy 781 MINIKFNLQCSVSLMNSMIPYGVKRLDDPDASIKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 783 MINIKFNLQCSVSLMNSMIPYGVKRLDDPDASIKDALLKYIYDNRGTLIGQVDRLLKDK 842
Qy 841 VNNLTSTDIPQLSKYVDNQRLSTFTTEYIK 871
Db 843 VNNLTSTDIPQLSKYVDNQRLSTFTTEYIK 873

RESULT 8

ADL90451
ID ADL90451 standard; protein; 873 AA.
XX ADL90451;
AC ADL90451;
XX
DT 17-JUN-2004 (first entry)
XX
DE Clostridial neurotoxin amino acid sequence SEQ ID NO:169.
XX
KW single chain polypeptide; clostridial neurotoxin light chain;
KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
KW botulinum; tetanus.
XX
OS Clostridium botulinum.
XX
FN WO2004024909-A2.
XX
PD 25-MAR-2004.
XX
FF 12-SEP-2003; 2003WO-GB003824.
XX
PR 12-SEP-2002; 2002US-00241596.
XX
PA (HEAL-) HEALTH PROTECTION AGENCY.
XX
FI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
FI Wayne J;
XX

DR WPI; 2004-270039/25.
XX N-PSDB; ADL90450.

PT New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.

PS Claim 1; SEQ ID NO 169; 588pp; English.

XX The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

SQ Sequence 873 AA;

Query Match 66.1%; Score 4509; DB 8; Length 873;

Best Local Similarity 99.8%; Pred. No. 9.1e-257;

Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNFQPNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHNKIWIPIERDFTTPEEGDLN 60

Db 3 MEFVNFQPNKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHNKIWIPIERDFTTPEEGDLN 62

Qy 61 PPPEAKQVPVSYDSTYLSLTDNEKNDYKGVTKLFEIYSTDGLRMLLTISVIRGIPFWGG 120

Db 63 PPPEAKQVPVSYDSTYLSLTDNEKNDYKGVTKLFEIYSTDGLRMLLTISVIRGIPFWGG 122

Qy 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIGPSADIIOFECKSFGEHVLNLTNGY 180

Db 123 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIGPSADIIOFECKSFGEHVLNLTNGY 182

Qy 181 GSTQYIRFSPDFTGFPESLEVDTPNLGAGKATDPAVTLAHELIYAGHRLVGIALNP 240

Db 183 GSTQYIRFSPDFTGFPESLEVDTPNLGAGKATDPAVTLAHELIYAGHRLVGIALNP 242

Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300

Db 243 RVFKVNTNAYEMSGLEVSFEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 302

Qy 301 KSIIVGTTASLQYMKNVKPEKYLLEDTSKGFSVDKLFEDKLYKMLTEIYTDNEDNFKPKV 360

Db 303 KSIIVGTTASLQYMKNVKPEKYLLEDTSKGFSVDKLFEDKLYKMLTEIYTDNEDNFKPKV 362

Qy 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQTEINNMFNFKLNFT 420

Db 363 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNGQTEINNMFNFKLNFT 422

Qy 421 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 480

Db 423 GLFEFYKLLCVRGIIITSKTSKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDNLKGE 482

Qy 481 ITSDNTIEAEBENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNIERFNG 540

XX

Db 483 ITSDTNIIEAEEENISLDLIQQYYLTFNFDNEPENISNIENSSDIIGOLELMPNIEFPNG 542
 Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKATEA 600
 Db 543 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKATEA 602
 Qy 601 AMELGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
 Db 603 AMELGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 662
 Qy 661 AVILLFPIEIPALVLTGTFALVSIAKVKLTQTVDNALSKRNEKWDDEVYKIVTNWLAK 720
 Db 663 AVILLFPIEIPALVLTGTFALVSIAKVKLTQTVDNALSKRNEKWDDEVYKIVTNWLAK 722
 Qy 721 VNTQIDILIRKMKKEALENQAETKAIINYQVQYTEEEKNNINFNIDDLSSKLNESINKA 780
 Db 723 VNTQIDILIRKMKKEALENQAETKAIINYQVQYTEEEKNNINFNIDDLSSKLNESINKA 782
 Qy 781 MININKFLNOCVSYLNMNMPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 840
 Db 783 MININKFLNOCVSYLNMNMPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 842
 Qy 841 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIK 871
 Db 843 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIK 873

RESULT 9
 AAWS6019
 ID AAWS6019 standard; protein; 871 AA.
 AC AAWS6019;
 XX
 XX 27-JUL-1998 (first entry)
 XX Recombinant botulinum neurotoxin type A LH423/A (Q2E,N26K,A27Y).
 DE Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
 KW Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen;
 XX detection; tetanus; non-toxic; toxin.
 OS Synthetic.
 OS Clostridium botulinum.
 XX
 XX WO9807864-A1.
 XX
 XX 26-FEB-1998.
 XX
 XX 22-AUG-1997; 97WO-GB0022273.
 XX
 XX 23-AUG-1996; 96GB-00017671.
 PR 13-DEC-1996; 96GB-00025996.
 XX
 XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PA (SPEY-) SPEYWOOD LAB LTD.
 XX
 XX Shone CC, Quinn CP, Foster KA;
 PI
 XX WPI; 1998-169168/15.
 DR N-PSDB; AAV26291.
 XX
 XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents,
 PT immunogens or as non-toxic standards for the detection of neurotoxins.
 XX
 XX Example 1; Page 108-111; 137pp; English.
 XX
 CC The present sequence represents a recombinant neurotoxin protein from the
 CC present invention. The present invention describes recombinant neurotoxin
 CC proteins which comprise a first and second domain, where the first domain
 CC is adapted to cleave one or more vesicle or plasma-membrane associated
 CC proteins essential to exocytosis, and where the second domain is adapted:
 CC (a) to translocate the protein into a cell; (b) to increase the
 CC solubility of the protein compared to the solubility of the first domain
 CC on its own, or (c) both to translocate the protein into a cell and to

CC increase the solubility of the protein compared to the solubility of the
 CC first domain on its own, the protein being free of clostridial neurotoxin
 CC (CN) and free of CN precursor that can be converted into toxin by
 CC proteolytic action. The recombinant proteins can be used as therapeutic
 CC agents for targeting cells expressing a relevant substrate. The products
 CC can also be used as immunogens and as non-toxic standards for the
 CC assessment and development of in vitro assays for the detection of
 CC functional botulinum or tetanus neurotoxins either in foodstuffs or in
 CC environmental samples
 XX
 SQ Sequence 871 AA;
 Query Match 66.1%; Score 4507; DB 2; Length 871;
 Best Local Similarity 99.8%; Pred. No. 1.2e-256;
 Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MPFVKQFNKDPVNGVDIAIYIKIPNAGOMQPVKAPKIHNKIWIPIERDFTFNPEEGDLN 60
 Db 1 MQFVNKQFNKDPVNGVDIAIYIKIPNAGOMQPVKAPKIHNKIWIPIERDFTFNPEEGDLN 60
 Qy 61 PPPEAKQVPVSYYDSTVLTSTDNKONKYLKGVTKLFEIYISTDLGRMLLTISVIRGIPFWGG 120
 Db 61 PPPEAKQVPVSYYDSTVLTSTDNKONKYLKGVTKLFEIYISTDLGRMLLTISVIRGIPFWGG 120
 Qy 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIGPSADIIOFECKSFGEHLVNLTRNGY 180
 Db 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIGPSADIIOFECKSFGEHLVNLTRNGY 180
 Qy 181 GSTQYIRFSPDFTGFEESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAINPN 240
 Db 181 GSTQYIRFSPDFTGFEESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAINPN 240
 Qy 241 RVFKVNTNAYENSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKFDAIASTLNKA 300
 Db 241 RVFKVNTNAYENSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKFDAIASTLNKA 300
 Qy 301 KSVIGTTASLOYMNKVPKEKYLISEDTSGKFSVDKLFKLYKMLTBIYTDNEDNPFVKFKV 360
 Db 301 KSVIGTTASLOYMNKVPKEKYLISEDTSGKFSVDKLFKLYKMLTBIYTDNEDNPFVKFKV 360
 Qy 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNLNTNLAANFNQNTNINNNFTKLKNFT 420
 Db 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNLNTNLAANFNQNTNINNNFTKLKNFT 420
 Qy 421 GLFEFYKLLCVRGIIITSKTSKLDKGNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
 Db 421 GLFEFYKLLCVRGIIITSKTSKLDKGNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
 Qy 481 ITSDTNIIEAEEENISLDLIQQYYLTFNFDNEPENISNIENSSDIIGOLELMPNIEFPNG 540
 Db 481 ITSDTNIIEAEEENISLDLIQQYYLTFNFDNEPENISNIENSSDIIGOLELMPNIEFPNG 540
 Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKATEA 600
 Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKVKATEA 600
 Qy 601 AMELGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
 Db 601 AMELGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
 Qy 661 AVILLFPIEIPALVLTGTFALVSIAKVKLTQTVDNALSKRNEKWDDEVYKIVTNWLAK 720
 Db 661 AVILLFPIEIPALVLTGTFALVSIAKVKLTQTVDNALSKRNEKWDDEVYKIVTNWLAK 720
 Qy 721 VNTQIDILIRKMKKEALENQAETKAIINYQVQYTEEEKNNINFNIDDLSSKLNESINKA 780
 Db 721 VNTQIDILIRKMKKEALENQAETKAIINYQVQYTEEEKNNINFNIDDLSSKLNESINKA 780
 Qy 781 MININKFLNOCVSYLNMNMPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 840
 Db 781 MININKFLNOCVSYLNMNMPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRDKDK 840
 Qy 841 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIK 871

25-MAR-2004.
 12-SEP-2003; 2003WO-GB003824.
 12-SEP-2002; 2002US-00241596.
 (HEAL-) HEALTH PROTECTION AGENCY.
 Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 Wayne J;
 WPI; 2004-270039/25.
 N-PSDB; ADL90434.
 New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for developing vaccines against clostridial toxin.
 Claim 1; SEQ ID NO 153; 588pp; English.

The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Query Match 66.1%; Score 4507; DB 8; Length 871;
 Best Local Similarity 99.8%; Pred. No. 1.2e-256;
 Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 MPFVNQFNKDPVNGVDIAIYKIPNAGQMPVKAFKIHKNKIWIPIRDTFTNPEGDNL 60
 1 MEFVNQFNKDPVNGVDIAIYKIPNAGQMPVKAFKIHKNKIWIPIRDTFTNPEGDNL 60
 61 PPEAKQVPVSYDYSTYLDNEKONYLKGVTLKERIYSTDLGRMLLTISVRGIPWGG 120
 61 PPEAKQVPVSYDYSTYLDNEKONYLKGVTLKERIYSTDLGRMLLTISVRGIPWGG 120
 121 STIDTELKVIDTNCINIVIPQSGYSRSEELNLIIGPSADIIQPECKSFGEVLNLTNGY 180
 121 STIDTELKVIDTNCINIVIPQSGYSRSEELNLIIGPSADIIQPECKSFGEVLNLTNGY 180
 181 GSTQYIRFSDFTFGPESLEVDNTNLLGAKGATDPAVTLAHELIVAGHRLYGIAPNP 240
 181 GSTQYIRFSDFTFGPESLEVDNTNLLGAKGATDPAVTLAHELIVAGHRLYGIAPNP 240
 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRIYYNKKFKDIASLANKA 300
 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRIYYNKKFKDIASLANKA 300
 301 KSIIVGTASIQYMNQVPEKYLISEDTSGKFSVDKLFKDKLYKMLTBIYTEDNFVKPKV 360
 301 KSIIVGTASIQYMNQVPEKYLISEDTSGKFSVDKLFKDKLYKMLTBIYTEDNFVKPKV 360

361 LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLNTLAANFNQONTNINNNFTKLNFT 420
 361 LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLNTLAANFNQONTNINNNFTKLNFT 420
 421 GUFEPYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
 421 GUFEPYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
 481 ITSNTNIEAAEENISLDLIQQYYLTNFNPEPENISIEIENLSSDIIIGOLELMPNIERPPNG 540
 481 ITSNTNIEAAEENISLDLIQQYYLTNFNPEPENISIEIENLSSDIIIGOLELMPNIERPPNG 540
 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVYTFPSSDYVKKVNKATEA 600
 541 KKYELDKYTMFHYLRAQEFHKGSRITATNSVNEALLNPSRVYTFPSSDYVKKVNKATEA 600
 601 AMPLGHWVEQLVVDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG 660
 601 AMPLGHWVEQLVVDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG 660
 661 AVILLEFIEIAPVLGTALVSYIANKVLTVTIDNALSKRNEKWDVYKYIVTNWLAK 720
 661 AVILLEFIEIAPVLGTALVSYIANKVLTVTIDNALSKRNEKWDVYKYIVTNWLAK 720
 721 VNTQIDILIRKKMKEALENQAETKAIINQYNOYTEEEKNNINFNIDDLSSKLNESINKA 780
 721 VNTQIDILIRKKMKEALENQAETKAIINQYNOYTEEEKNNINFNIDDLSSKLNESINKA 780
 781 MININKFLNQCYSYLMNSMIPYGVKRLBDFDASLKDALLKYIDNRTGLIGQVDRLLKDK 840
 781 MININKFLNQCYSYLMNSMIPYGVKRLBDFDASLKDALLKYIDNRTGLIGQVDRLLKDK 840
 841 VNNLTSTDIPFQLSKVVDNORLLSTFTTEYIK 871
 841 VNNLTSTDIPFQLSKVVDNORLLSTFTTEYIK 871

RESULT 12
 ADL90308
 ID ADL90308 standard; protein; 871 AA.
 XX ADL90308;
 AC ADL90308;
 DT 17-JUN-2004 (first entry)
 XX C. botulinum LH423/A (Q2E,N26K,A27V) protein SEQ ID NO:26.
 KW single chain polypeptide; clostridial neurotoxin light chain;
 KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
 KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
 KW botulinum; tetanus.
 OS Clostridium botulinum.
 XX WO2004024909-A2.
 PN 25-MAR-2004.
 PD 12-SEP-2003; 2003WO-GB003824.
 PF 12-SEP-2002; 2002US-00241596.
 PR (HEAL-) HEALTH PROTECTION AGENCY.
 XX Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 PI Wayne J;
 XX WPI; 2004-270039/25.
 DR N-PSDB; ADL90307.
 DR New single chain polypeptides comprising clostridial neurotoxin light and heavy chains, useful as positive controls for toxin assays, or for

developing vaccines against clostridial toxin.

Example 1; SEQ ID NO 26; 588pp; English.

The present invention describes a single chain polypeptide comprising clostridial neurotoxin light and heavy chains. The single chain polypeptide comprises 2 domains: the first domain is a clostridial neurotoxin light chain, or its fragment or variant, which is capable of cleaving one or more vesicle or plasma membrane associated proteins essential to exocytosis; the second domain is a clostridial neurotoxin heavy chain H-N portion, or its fragment or variant, which is capable of translocating the polypeptide into a cell and/or increasing the solubility of the polypeptide compared to the solubility of the first domain on its own. The second domain lacks a functional C-terminal part of a clostridial neurotoxin heavy chain, designated H-C, which renders the polypeptide incapable of binding to cell surface receptors that are the natural cell surface receptors to which native clostridial neurotoxin binds. Also described is a nucleic acid molecule encoding the single chain polypeptide described above. The single chain polypeptide has antibacterial activity, and can be used in vaccines. The single chain polypeptides can be used as positive controls for toxin assays, as reagent components for the synthesis of therapeutic molecules, or for developing vaccines against clostridial toxin. The polypeptides are also useful as non-toxic standards for the assessment and development of in vitro assays for detection of functional botulinum or tetanus neurotoxins in foodstuffs or environmental samples. The present sequence is used in the exemplification of the present invention.

Sequence 871 AA;

Query Match 66.1%; Score 4507; DB 8; Length 871;
Best Local Similarity 99.8%; Pred. No. 1.2e-256;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPFVKNQFNVKDPVNGVDIAYIKIPNAGQVPKAFKHINKIWIIPERDFTTPEEGDLN 60
DB 1 MQFVKNQFNVKDPVNGVDIAYIKIPNAGQVPKAFKHINKIWIIPERDFTTPEEGDLN 60
QY 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTFLFERIYSTDLGRMLTISVIRGIPFWGG 120
DB 61 PPPEAKQVPVSYDSTYLSSTNEKNDYLGKVTFLFERIYSTDLGRMLTISVIRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQPDGVSRSBELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQPDGVSRSBELNLVIIGPSADIIQFECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKFAFDPAVTLAHLIYAGHRLYGIANPN 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKFAFDPAVTLAHLIYAGHRLYGIANPN 240
QY 241 RVFKVNTNAYVMSGLVSPFELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYVMSGLVSPFELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
QY 301 KSIIVGTTASLQYMNKVFKEKYLISEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
DB 301 KSIIVGTTASLQYMNKVFKEKYLISEDTSKGFSVDKLFKDKLYKMLTEIYTEDNPFVKFKV 360
QY 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNFLRNLTNLAANFGONTINNNNFTKLKNFT 420
DB 361 LNRKTYLNFDAVKFKNIVPKVNTIYDGNFLRNLTNLAANFGONTINNNNFTKLKNFT 420
QY 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPEDNFTNDLNKGE 480
DB 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPEDNFTNDLNKGE 480
QY 481 ITSDNTIEAABENISLDLIQOYYLTFNFDNPEPENISIENTIENSSDIIGOLELMPNIERP 540
DB 481 ITSDNTIEAABENISLDLIQOYYLTFNFDNPEPENISIENTIENSSDIIGOLELMPNIERP 540
QY 541 KKYELDKYTFHLYRAQEFHGHKGRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTFHLYRAQEFHGHKGRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600

QY 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660
QY 661 AVILLEPIPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDDEVKYIIVTNWLAK 720
DB 661 AVILLEPIPIAIPVLGTFFALVSIANKVLTQVOTIDNALSKRNEKWDDEVKYIIVTNWLAK 720
QY 721 VNTQIDILIRKMKAELENOAEATKAIINYOVNVTREKKNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKMKAELENOAEATKAIINYOVNVTREKKNINFNIDDLSSKLNESINKA 780
QY 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
DB 781 MININKFLNOCSSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
QY 841 VNNLTSTDIPFQPSKYVDNORLLSTFTTEYIK 871
DB 841 VNNLTSTDIPFQPSKYVDNORLLSTFTTEYIK 871

RESULT 13

ADL90284
ID ADL90284 standard; protein; 871 AA.

XX AC ADL90284;

XX DT 17-JUN-2004 (first entry)

XX DE Clostridium botulinum LH423/A protein SEQ ID NO:2.

XX KW single chain polypeptide; clostridial neurotoxin light chain;
XX KW clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
XX KW antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX KW botulinum; tetanus.

XX OS Clostridium botulinum.

XX PN WO2004024909-A2.

XX PD 25-MAR-2004.

XX PF 12-SEP-2003; 2003WO-GB003824.

XX PR 12-SEP-2002; 2002US-00241596.

XX PA (HEAL-) HEALTH PROTECTION AGENCY.

XX PI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
XX PI Wayne J;

XX XX WPI; 2004-270039/25.

XX DR N-PSDB; ADL90283.

XX XX New single chain polypeptides comprising clostridial neurotoxin light and
XX PT heavy chains, useful as positive controls for toxin assays, or for
XX PT developing vaccines against clostridial toxin.

XX PS Example 1; SEQ ID NO 2; 588pp; English.

XX CC The present invention describes a single chain polypeptide comprising
XX CC clostridial neurotoxin light and heavy chains. The single chain
XX CC polypeptide comprises 2 domains: the first domain is a clostridial
XX CC neurotoxin light chain, or its fragment or variant, which is capable of
XX CC cleaving one or more vesicle or plasma membrane associated proteins
XX CC essential to exocytosis; the second domain is a clostridial neurotoxin
XX CC heavy chain H-N portion, or its fragment or variant, which is capable of
XX CC translocating the polypeptide into a cell and/or increasing the
XX CC solubility of the polypeptide compared to the solubility of the first
XX CC domain on its own. The second domain lacks a functional C-terminal part
XX CC of a clostridial neurotoxin heavy chain, designated H-C, which renders
XX CC the polypeptide incapable of binding to cell surface receptors that are

CC the natural cell surface receptors to which native clostridial neurotoxin
 CC binds. Also described is a nucleic acid molecule encoding the single
 CC chain polypeptide described above. The single chain polypeptide has
 CC antibacterial activity, and can be used in vaccines. The single chain
 CC polypeptides can be used as positive controls for toxin assays, as
 CC reagent components for the synthesis of therapeutic molecules, or for
 CC developing vaccines against clostridial toxin. The polypeptides are also
 CC useful as non-toxic standards for the assessment and development of in
 CC vitro assays for detection of functional botulinum or tetanus neurotoxins
 CC in foodstuffs or environmental samples. The present sequence is used in
 CC the exemplification of the present invention.

XX Sequence 871 AA;

Query Match 66.1%; Score 4507; DB 8; Length 871;

Best Local Similarity 99.8%; Pred. No. 1.2e-256;

Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNKQPNYKDPVNGVDIAIYIKIPNAGQMPVKAFAKHINKIWIPIERDFTFNPBEGDLN 60
 Db 1 MQFVNKQPNYKDPVNGVDIAIYIKIPNAGQMPVKAFAKHINKIWIPIERDFTFNPBEGDLN 60
 Qy 61 PPPEAKQVPVSYDYSTYLSDTNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
 Db 61 PPPEAKQVPVSYDYSTYLSDTNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
 Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQPECKSPGHEVLNLTNGY 180
 Db 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIQPECKSPGHEVLNLTNGY 180
 Qy 181 GSTQYIRFSDPFTGPFESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIAPN 240
 Db 181 GSTQYIRFSDPFTGPFESLEVDNPLLGAGKATDPAVTLAHLIIYAGHRLYGIAPN 240
 Qy 241 RVKQNTNAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA 300
 Db 241 RVKQNTNAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA 300
 Qy 301 KSIIVGTTASLQYMNKVFKEKYLLEDSTSGKFSVDKLPDKLYKMLTEIYTEDNFVKPFV 360
 Db 301 KSIIVGTTASLQYMNKVFKEKYLLEDSTSGKFSVDKLPDKLYKMLTEIYTEDNFVKPFV 360
 Qy 361 LNRKTYLNFDFKAVKINIVPKVNTIYDGNFLRNLTNLAANFNGQNTINNMFNFKLNFT 420
 Db 361 LNRKTYLNFDFKAVKINIVPKVNTIYDGNFLRNLTNLAANFNGQNTINNMFNFKLNFT 420
 Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYKALNDLCIKVNNWDLRPSSEDFNTDNLNGEE 480
 Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYKALNDLCIKVNNWDLRPSSEDFNTDNLNGEE 480
 Qy 481 ITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISIEINLSSDIIGOLELMPNIERFPNG 540
 Db 481 ITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISIEINLSSDIIGOLELMPNIERFPNG 540
 Qy 541 KKYELDKYTFHLYRAQEFHGKGRIALTNSVNEALLNPRVYTFSSDYVKKVKNKATEA 600
 Db 541 KKYELDKYTFHLYRAQEFHGKGRIALTNSVNEALLNPRVYTFSSDYVKKVKNKATEA 600
 Qy 601 AMFLGWVEQLVYDFTDSTSEVTTDKIADITIIIPYIGPALNIGMLYKDDFVGCALIFSG 660
 Db 601 AMFLGWVEQLVYDFTDSTSEVTTDKIADITIIIPYIGPALNIGMLYKDDFVGCALIFSG 660
 Qy 661 AVILLEPIEIPAIVLGTFFALVSIVANKVLTQVOTIDNALSKRNEKWDVYKIVTNWLAK 720
 Db 661 AVILLEPIEIPAIVLGTFFALVSIVANKVLTQVOTIDNALSKRNEKWDVYKIVTNWLAK 720
 Qy 721 VNTQIDILIRKQKALENQAETKAIINYQNVQTEBEKKNNINFDLSSKLNESINKA 780
 Db 721 VNTQIDILIRKQKALENQAETKAIINYQNVQTEBEKKNNINFDLSSKLNESINKA 780
 Qy 781 MININKFLNOCSSVLYMNSMIPYGVKRLDFDASIKDALLKYIYDNRGTLIGQVDRUKDK 840
 Db 781 MININKFLNOCSSVLYMNSMIPYGVKRLDFDASIKDALLKYIYDNRGTLIGQVDRUKDK 840

Qy 841 VNNVLTSTIDIPQLSKYVDNORLLSTFTTEYIK 871
 Db 841 VNNVLTSTIDIPQLSKYVDNORLLSTFTTEYIK 871

RESULT 14

ADL90433
 ID ADL90433 standard; protein; 873 AA.

XX AC ADL90433;

DT 17-JUN-2004 (first entry)

XX Clostridial neurotoxin amino acid sequence SEQ ID NO:151.

XX single chain polypeptide; clostridial neurotoxin light chain;
 XX clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
 XX antibacterial; vaccine; toxin assay; clostridial toxin; detection;
 XX Clostridium botulinum.

OS Clostridium botulinum.

PN W02004024909-A2.

XX 25-MAR-2004.

PF 12-SEP-2003; 2003WO-GB003824.

PR 12-SEP-2002; 2002US-00241596.

XX (HEAL-) HEALTH PROTECTION AGENCY.

PI Shone CC, Foster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
 PI Wayne J;

DR WPI; 2004-270039/25.

XX N-PSDB; ADL90432.

XX New single chain polypeptides comprising clostridial neurotoxin light and
 XX heavy chains, useful as positive controls for toxin assays, or for
 XX developing vaccines against clostridial toxin.

PS Claim 1; SEQ ID NO 151; 588pp; English.

XX The present invention describes a single chain polypeptide comprising
 CC clostridial neurotoxin light and heavy chains. The single chain
 CC polypeptide comprises 2 domains: the first domain is a clostridial
 CC neurotoxin light chain, or its fragment or variant, which is capable of
 CC cleaving one or more vesicle or plasma membrane associated proteins
 CC essential to exocytosis; the second domain is a clostridial neurotoxin
 CC heavy chain H-N portion, or its fragment or variant, which is capable of
 CC translocating the polypeptide into a cell and/or increasing the
 CC solubility of the polypeptide compared to the solubility of the first
 CC domain on its own. The second domain lacks a functional C-terminal part
 CC of a clostridial neurotoxin heavy chain, designated H-C, which renders
 CC the polypeptide incapable of binding to cell surface receptors that are
 CC the natural cell surface receptors to which native clostridial neurotoxin
 CC binds. Also described is a nucleic acid molecule encoding the single
 CC chain polypeptide described above. The single chain polypeptide has
 CC antibacterial activity, and can be used in vaccines. The single chain
 CC polypeptides can be used as positive controls for toxin assays, as
 CC reagent components for the synthesis of therapeutic molecules, or for
 CC developing vaccines against clostridial toxin. The polypeptides are also
 CC useful as non-toxic standards for the assessment and development of in
 CC vitro assays for detection of functional botulinum or tetanus neurotoxins
 CC in foodstuffs or environmental samples. The present sequence is used in
 CC the exemplification of the present invention.

XX Sequence 873 AA;

Query Match 66.1%; Score 4507; DB 8; Length 873;

Best Local Similarity 99.8%; Pred. No. 1.2e-256;

Matches		869;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	MPFVNKQFNKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIPIERDFTFTPEEGDLN	60							
Db	3	MEFVNKQFNKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIPIERDFTFTPEEGDLN	62							
Qy	61	PPPEAKQVPVSYDSTYLSSTDNKDKYLGKVTGKLFERIVSTDLGRMLLTSIVRGIPFWGG	120							
Db	63	PPPEAKQVPVSYDSTYLSSTDNKDKYLGKVTGKLFERIVSTDLGRMLLTSIVRGIPFWGG	122							
Qy	121	STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	180							
Db	123	STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY	182							
Qy	181	GSTQYIRFSDPFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAINPN	240							
Db	183	GSTQYIRFSDPFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAINPN	242							
Qy	241	RVFKVNTNAYENSGLEVSFEELRTFGCHDAKFTDSIQENEFRLYYNKKFKDIASLTNKA	300							
Db	243	RVFKVNTNAYENSGLEVSFEELRTFGCHDAKFTDSIQENEFRLYYNKKFKDIASLTNKA	302							
Qy	301	KSIIVGTTASIQYKMNVPKPKYLLSDETSKGFSVDKLFKLYKMLTEIYTEDNFVFPKV	360							
Db	303	KSIIVGTTASIQYKMNVPKPKYLLSDETSKGFSVDKLFKLYKMLTEIYTEDNFVFPKV	362							
Qy	361	LNKTYLNFDAKVPKINIVPKVNTIYDGFNLNLTNLAANFNPGONTNINNNFTKLKQFT	420							
Db	363	LNKTYLNFDAKVPKINIVPKVNTIYDGFNLNLTNLAANFNPGONTNINNNFTKLKQFT	422							
Qy	421	GLFFFYKLLCVRGIIITSKTSKSLDKGVNKAINDLCIKVNNWDLFPSPSEDNFTDLNKGEE	480							
Db	423	GLFFFYKLLCVRGIIITSKTSKSLDKGVNKAINDLCIKVNNWDLFPSPSEDNFTDLNKGEE	482							
Qy	481	ITSDTNEAABENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNIERPNG	540							
Db	483	ITSDTNEAABENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNIERPNG	542							
Qy	541	KYELDKYTFHYLRADQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600							
Db	543	KYELDKYTFHYLRADQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	602							
Qy	601	AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG	660							
Db	603	AMFLGWELQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG	662							
Qy	661	AVILLEFPIETAIPIVGLTFALVSYANKVLTQVTDNALSKRNEKWDVYKVIYTNWLAK	720							
Db	663	AVILLEFPIETAIPIVGLTFALVSYANKVLTQVTDNALSKRNEKWDVYKVIYTNWLAK	722							
Qy	721	VNTQIDLRKWKKEALENOAETKAIINYOYNQYTEBEKNNINFNIDLSKLNESINKA	780							
Db	723	VNTQIDLRKWKKEALENOAETKAIINYOYNQYTEBEKNNINFNIDLSKLNESINKA	782							
Qy	781	MININKFNQCSVSLMNSMIPYGVKRLDFTDASLKDALIKYIYDNRGTLIGQVDRLKDK	840							
Db	783	MININKFNQCSVSLMNSMIPYGVKRLDFTDASLKDALIKYIYDNRGTLIGQVDRLKDK	842							
Qy	841	VNNTLSTDIPFQLSKYVDNQRLSTFTTEYIK	871							
Db	843	VNNTLSTDIPFQLSKYVDNQRLSTFTTEYIK	873							
RESULT 15										
ID	ADL90431	standard; protein; 873 AA.								
XX	AC	ADL90431;								
XX	DT	17-JUN-2004 (first entry)								
XX	DE	Clostridial neurotoxin amino acid sequence SEQ ID NO:149.								

KW	single chain polypeptide; clostridial neurotoxin light chain;
KW	clostridial neurotoxin heavy chain; Clostridium; neurotoxin; exocytosis;
KW	antibacterial; vaccine; toxin assay; clostridial toxin; detection;
XX	botulinum; tetanus.
OS	Clostridium botulinum.
XX	WO2004024909-A2.
PN	25-MAR-2004.
XX	12-SEP-2003; 2003WO-GB003824.
PF	12-SEP-2002; 2002US-00241596.
XX	(HEAL-) HEALTH PROTECTION AGENCY.
XX	Shone CC, Poster KA, Chaddock J, Marks P, Sutton MJ, Stancombe P;
PI	Wayne J;
PI	WPI; 2004-270039/25.
XX	N-PSDB; ADL90430.
DR	New single chain polypeptides comprising clostridial neurotoxin light and
XX	heavy chains, useful as positive controls for toxin assays, or for
PT	developing vaccines against clostridial toxin.
PT	Claim 1; SEQ ID NO 149; 588pp; English.
XX	The present invention describes a single chain polypeptide comprising
XX	clostridial neurotoxin light and heavy chains. The single chain
CC	polypeptide comprises 2 domains: the first domain is a clostridial
CC	neurotoxin light chain, or its fragment or variant, which is capable of
CC	cleaving one or more vesicle or plasma membrane associated proteins
CC	essential to exocytosis; the second domain is a clostridial neurotoxin
CC	heavy chain H-N portion, or its fragment or variant, which is capable of
CC	translocating the polypeptide into a cell and/or increasing the
CC	solubility of the polypeptide compared to the solubility of the first
CC	domain on its own. The second domain lacks a functional C-terminal part
CC	of a clostridial neurotoxin heavy chain, designated H-C, which renders
CC	the polypeptide incapable of binding to cell surface receptors that are
CC	the natural cell surface receptors to which native clostridial neurotoxin
CC	binds. Also described is a nucleic acid molecule encoding the single
CC	chain polypeptide described above. The single chain polypeptide has
CC	antibacterial activity, and can be used in vaccines. The single chain
CC	polypeptides can be used as positive controls for toxin assays, as
CC	reagent components for the synthesis of therapeutic molecules, or for
CC	developing vaccines against clostridial toxin. The polypeptides are also
CC	useful as non-toxic standards for the assessment and development of in
CC	vitro assays for detection of functional botulinum or tetanus neurotoxins
CC	in foodstuffs or environmental samples. The present sequence is used in
CC	the exemplification of the present invention.
XX	Sequence 873 AA;
SQ	Query Match 66.1%; Score 4507; DB 8; Length 873;
	Best Local Similarity 99.8%; Pred. No. 1.2e-256;
	Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy	1 MPFVNKQFNKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIPIERDFTFTPEEGDLN 60
Db	3 MEFVNKQFNKDPVNGVDIAIYIKIPNAGQMPKAFKIHKNKIWIPIERDFTFTPEEGDLN 62
Qy	61 PPPEAKQVPVSYDSTYLSSTDNKDKYLGKVTGKLFERIVSTDLGRMLLTSIVRGIPFWGG 120
Db	63 PPPEAKQVPVSYDSTYLSSTDNKDKYLGKVTGKLFERIVSTDLGRMLLTSIVRGIPFWGG 122
Qy	121 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 180
Db	123 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQFECKSFGEHVLNLTNGY 182
Qy	181 GSTQYIRFSDPFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAINPN 240

Db	183	GSTQYIRFSDFTFGFESLEVDNPLLGAGKFAFDPAVTLAHELHAGHRLYGIAINPN	242
Qy	241	RVFKVNTNAYEMSGLEVSPFELRTFGGHDAKFTIDSLOENEFRLYYNKKFKDIASTLNKA	300
Db	243	RVFKVNTNAYEMSGLEVSPFELRTFGGHDAKFTIDSLOENEFRLYYNKKFKDIASTLNKA	302
Qy	301	KSIVGTTASLOYMKQNVFKYKLLSEDSGKFSVDKLPDKLYKMLTEIYTEDNPFVKPKV	360
Db	303	KSIVGTTASLOYMKQNVFKYKLLSEDSGKFSVDKLPDKLYKMLTEIYTEDNPFVKPKV	362
Qy	361	LNKTYLNFDPKAVFKINIYDGFNLRNTNLAANFNGONTENNMMFTKLKNFT	420
Db	363	LNKTYLNFDPKAVFKINIYDGFNLRNTNLAANFNGONTENNMMFTKLKNFT	422
Qy	421	GLPEFYKLLCVRGIIITSKTSLODKYGNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE	480
Db	423	GLPEFYKLLCVRGIIITSKTSLODKYGNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGE	482
Qy	481	ITSDTNIEAAEENISLDLIQOYLYTFNFDNEPENISIENTLSSDIIGOLELMPNIERFPNG	540
Db	483	ITSDTNIEAAEENISLDLIQOYLYTFNFDNEPENISIENTLSSDIIGOLELMPNIERFPNG	542
Qy	541	KYELDKYTFPHYLRAOEFEGHKSRIALTNSVNEALLNPSRVYTFPSSDYVKKVNKATEA	600
Db	543	KYELDKYTFPHYLRAOEFEGHKSRIALTNSVNEALLNPSRVYTFPSSDYVKKVNKATEA	602
Qy	601	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	660
Db	603	AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	662
Qy	661	AVILLEPIPEIAPVLGTFALVSVIANKVLTVQTDIDNALSKRNEKWDDEVYKYIVTNWLAK	720
Db	663	AVILLEPIPEIAPVLGTFALVSVIANKVLTVQTDIDNALSKRNEKWDDEVYKYIVTNWLAK	722
Qy	721	VNTQIDILIRKKMEALENQAETKAIINYQYNOYTEEEKNNINFNIDDLSSKLNESINKA	780
Db	723	VNTQIDILIRKKMEALENQAETKAIINYQYNOYTEEEKNNINFNIDDLSSKLNESINKA	782
Qy	781	MININKFLNOCVSVYLMNSMIPYGVKREDFDASLKDALLKYIYDNRGTILIGQVDRLKDK	840
Db	783	MININKFLNOCVSVYLMNSMIPYGVKREDFDASLKDALLKYIYDNRGTILIGQVDRLKDK	842
Qy	841	VNNTLSTDIPFQLSKYVDNQRLLSTFTTEYIK	871
Db	843	VNNTLSTDIPFQLSKYVDNQRLLSTFTTEYIK	873

Search completed: January 31, 2005, 14:04:11
Job time : 182 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 43 Seconds
(without alignments)
2008.047 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFNQFNKDPNGVDIA.....EFIPVDGCGERPLHHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6756	99.0	1296	1	US-08-480-604A-28
2	6756	99.0	1296	2	US-08-405-496A-28
3	6756	99.0	1296	3	US-08-915-136-28
4	6756	99.0	1296	4	US-09-084-517-28
5	4515	66.2	871	4	US-09-255-829-8
6	4507	66.1	871	4	US-09-255-829-2
7	4507	66.1	871	4	US-09-255-829-26
8	4501	66.0	1013	4	US-09-255-829-18
9	4495	65.9	873	4	US-09-255-829-6
10	4495	65.9	875	4	US-09-255-829-10
11	4495	65.9	894	4	US-09-255-829-4
12	4494.5	65.9	907	4	US-09-255-829-16
13	4494.5	65.9	953	4	US-09-255-829-14
14	4493.5	65.9	878	4	US-09-255-829-12
15	2348	34.4	448	4	US-09-288-326A-7
16	2307	33.8	438	1	US-08-480-604A-23
17	2307	33.8	438	2	US-08-405-496A-23
18	2307	33.8	438	3	US-08-915-136-23
19	2307	33.8	438	4	US-09-084-517-23
20	2307	33.8	462	1	US-08-480-604A-26
21	2307	33.8	462	2	US-08-405-496A-26
22	2307	33.8	462	3	US-08-915-136-26
23	2307	33.8	462	4	US-09-084-517-26
24	2167	31.8	423	4	US-09-288-326A-8
25	2146.5	31.5	1169	4	US-09-255-829-20
26	2036	29.8	382	4	US-09-288-326A-9
27	1892.5	27.7	1315	3	US-08-913-880C-1

28 1451.5 21.3 858 4 US-09-255-829-22 Sequence 22, Appl
29 1451.5 21.3 858 4 US-09-255-829-29 Sequence 23, Appl
30 1348 19.8 874 3 US-08-913-880C-11 Sequence 11, Appl
31 1348 19.8 875 3 US-08-913-880C-10 Sequence 10, Appl
32 1345 19.7 862 3 US-08-913-880C-14 Sequence 14, Appl
33 1345 19.7 865 3 US-08-913-880C-13 Sequence 13, Appl
34 1345 19.7 866 3 US-08-913-880C-12 Sequence 12, Appl
35 1343 19.7 853 3 US-08-913-880C-17 Sequence 17, Appl
36 1343 19.7 858 3 US-08-913-880C-16 Sequence 16, Appl
37 1343 19.7 860 3 US-08-913-880C-15 Sequence 15, Appl
38 1071 15.7 207 3 US-08-446-114A-21 Sequence 21, Appl
39 733 10.7 141 4 US-09-465-276-1 Sequence 1, Appl
40 727 10.7 140 3 US-08-446-114A-22 Sequence 22, Appl
41 653.5 9.6 618 1 US-08-668-381A-5 Sequence 5, Appl
42 646.5 9.5 548 4 US-09-255-829-24 Sequence 24, Appl
43 634.5 9.3 452 1 US-07-618-312A-2 Sequence 2, Appl
44 634.5 9.3 452 1 US-08-280-228-2 Sequence 2, Appl
45 633.5 9.3 452 1 US-07-618-312A-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-28
; Sequence 28, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-28

Query Match 99.0%; Score 6756; DB 2; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVKQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
DB 1 MQFVKQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60

QY 61 PPPEAKQPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQPVSYDSTYSLTDNEKDNKYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120

QY 121 STIDTELKVIDTNCINVIQDGSYSRSEELNLIIGPSADIIQPECKSPFGHEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYSRSEELNLIIGPSADIIQPECKSPFGHEVLNLTNGY 180

QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240

QY 241 RVFKVNTNAYEMSGLEVSFEELTFGHDADKFDISLOENEFRLYYNKKFKDIASLTNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELTFGHDADKFDISLOENEFRLYYNKKFKDIASLTNKA 300

QY 301 KSIQVGTASLOYKQNVKFKYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKFKV 360
DB 301 KSIQVGTASLOYKQNVKFKYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKFKV 360

QY 361 LNRKTYLNFKAQVFKINIPKVNVTIYDGNFLNLTNLAANFNGONTINNNTFKLNFT 420
DB 361 LNRKTYLNFKAQVFKINIPKVNVTIYDGNFLNLTNLAANFNGONTINNNTFKLNFT 420

QY 421 GLPEFYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
DB 421 GLPEFYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480

QY 481 ITSDTNTAEAEENISLDLIQOYLYTFNPDNEPENISNIENSSDIIGQELMPNIERPNG 540
DB 481 ITSDTNTAEAEENISLDLIQOYLYTFNPDNEPENISNIENSSDIIGQELMPNIERPNG 540

QY 541 KYELDKYTFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KYELDKYTFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600

QY 601 AMFLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660

QY 661 AVILLEPETAIPVLGTFAVSVIANKVLTVQIDNALSKEKKEWDEVKYIIVTNWLAK 720
DB 661 AVILLEPETAIPVLGTFAVSVIANKVLTVQIDNALSKEKKEWDEVKYIIVTNWLAK 720

QY 721 VNTQIDILIRKQWKEALNQAEATKAIINYQVQYTEBEKKNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKQWKEALNQAEATKAIINYQVQYTEBEKKNINFNIDDLSSKLNESINKA 780

QY 781 MININKFLNQCVSYLNMNMPYGVGRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK 840
DB 781 MININKFLNQCVSYLNMNMPYGVGRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK 840

QY 841 VNTLSTDPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900

DB 841 VNTLSTDPQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
QY 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTSFWIRIPKYNFNSISLNN 960
DB 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVNMYENFSTSFWIRIPKYNFNSISLNN 960

QY 961 EYTIINCMMNNSGKVSILNYGEIITWLODQEIQRVVFYKYSQMINISDYINRWIFVTIT 1020
DB 961 EYTIINCMMNNSGKVSILNYGEIITWLODQEIQRVVFYKYSQMINISDYINRWIFVTIT 1020

QY 1021 NNLNNSKIYINGRLIDOKPISNLGNIHASNIMFKLDGCRDTHRYIWIYKYNLFDKELN 1080
DB 1021 NNLNNSKIYINGRLIDOKPISNLGNIHASNIMFKLDGCRDTHRYIWIYKYNLFDKELN 1080

QY 1081 EKEIKOLDYDQNSGILKDFWGDYLVQDKPYMLNLYDPNKKYVDVNNVGRGYMLKGP 1140
DB 1081 EKEIKOLDYDQNSGILKDFWGDYLVQDKPYMLNLYDPNKKYVDVNNVGRGYMLKGP 1140

QY 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKONIVRNDRVYINVVKNKESYRLATNASQA 1200
DB 1141 GSVMTTNIYLNSSLYRGTKFIIKKYASGNKONIVRNDRVYINVVKNKESYRLATNASQA 1200

QY 1201 GVEKILSALEIPDVGNLSQVVMKSNDOQITNKKCNLQDNNNDIGFTGFHOFNNIAK 1260
DB 1201 GVEKILSALEIPDVGNLSQVVMKSNDOQITNKKCNLQDNNNDIGFTGFHOFNNIAK 1260

QY 1261 LVASNWNRQIERSSRTLGCSEWFIIPVDDGWMGRPL 1296
DB 1261 LVASNWNRQIERSSRTLGCSEWFIIPVDDGWMGRPL 1296

RESULT 3

US-08-915-136-28
; Sequence 28, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHAYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPND-01610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-517-28

Query Match 99.0%; Score 6756; DB 4; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVKQFNKDPVNGVDIAIKIPNAGQMPVKAFKIHKKIWIPIPERDFTNPEEGDLN 60
DB 1 MQFVKQFNKDPVNGVDIAIKIPNAGQMPVKAFKIHKKIWIPIPERDFTNPEEGDLN 60
QY 61 PPPEAKQVPVSYDYSTYLSLTDNEKDYLGKVTGLFERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQVPVSYDYSTYLSLTDNEKDYLGKVTGLFERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STDTTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
DB 121 STDTTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRPDPFTFGFESLEVDNPLLGAGKPAFDPAVLAHELIIYAGHRLYGIALNP 240
DB 181 GSTQYIRPDPFTFGFESLEVDNPLLGAGKPAFDPAVLAHELIIYAGHRLYGIALNP 240
QY 241 RVFKVNTNAYEMSGLEVSFEELTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELTFGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300
QY 301 KSIIVGTTASLOYMKNVPEKYLLEDTSFGKPSVDKLFKLYKMLTEIYTEDNPFVKPKV 360
DB 301 KSIIVGTTASLOYMKNVPEKYLLEDTSFGKPSVDKLFKLYKMLTEIYTEDNPFVKPKV 360
QY 361 LNRKTYLNFDAVKFKINIVPKVNTIYDGFNLRLNLAANPFGONTENNMFKLKNET 420
DB 361 LNRKTYLNFDAVKFKINIVPKVNTIYDGFNLRLNLAANPFGONTENNMFKLKNET 420
QY 421 GLFEFYKLLCVRGIIITKTSKSLDKYKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
DB 421 GLFEFYKLLCVRGIIITKTSKSLDKYKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
QY 481 ITSDNTIEAENISLDLIQOYIYTFNPDNENISIENTSSDIIGOLEMPNTERPFG 540
DB 481 ITSDNTIEAENISLDLIQOYIYTFNPDNENISIENTSSDIIGOLEMPNTERPFG 540
QY 541 KKYELDKYTFMPLYRAQEPFEGKGRIALTNSVNEALLNPSRVYTFPSSDYVKKYKATEA 600
DB 541 KKYELDKYTFMPLYRAQEPFEGKGRIALTNSVNEALLNPSRVYTFPSSDYVKKYKATEA 600
QY 601 AMFLGWELQVYDFTDTSVSTTDKADIITIIPIYGPALNIGMLYKDDFVGLIFSG 660
DB 601 AMFLGWELQVYDFTDTSVSTTDKADIITIIPIYGPALNIGMLYKDDFVGLIFSG 660
QY 661 AVILLEFIPEIAPVLGTGTFALVSIANKVLTQVOTIDNLSKRNEKWDDEVYKIVTNWLAK 720
DB 661 AVILLEFIPEIAPVLGTGTFALVSIANKVLTQVOTIDNLSKRNEKWDDEVYKIVTNWLAK 720

661 AVILLEFIPEIAPVLGTGTFALVSIANKVLTQVOTIDNLSKRNEKWDDEVYKIVTNWLAK 720
QY 721 VNTQIDILIRKKMKEALENQAEATKAIINYQYNTTEEKNNINFPNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKKMKEALENQAEATKAIINYQYNTTEEKNNINFPNIDDLSSKLNESINKA 780
QY 781 MINIKFLAQCSVSYLMNSMIPYGVKRLDPFASLKDALKIYDNRGTGILIGQVDRUKDK 840
DB 781 MINIKFLAQCSVSYLMNSMIPYGVKRLDPFASLKDALKIYDNRGTGILIGQVDRUKDK 840
QY 841 VNNLTSTDIPIFOLSKYVDNORLLSTTEYIKNIINTSILNLRYESNHLIDLSRVASKINI 900
DB 841 VNNLTSTDIPIFOLSKYVDNORLLSTTEYIKNIINTSILNLRYESNHLIDLSRVASKINI 900
QY 901 GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYFNISILNN 960
DB 901 GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNMYENFSTFWIRIPKYFNISILNN 960
QY 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWFVTIT 1020
DB 961 EYTIINCMMNSGKVSILNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWFVTIT 1020
QY 1021 NNRLLNSKIYINGRLIDOKPISNLGNHASNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
DB 1021 NNRLLNSKIYINGRLIDOKPISNLGNHASNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
QY 1081 EKEIKOLYDNQNSGILKDFWGDYLYQDKFYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
DB 1081 EKEIKOLYDNQNSGILKDFWGDYLYQDKFYMLNLYDPNKYVDVNNVGIRGYMYLKGPR 1140
QY 1141 GSVMTTNIYNSLSYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVQKRYRLATNASQA 1200
DB 1141 GSVMTTNIYNSLSYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVQKRYRLATNASQA 1200
QY 1201 GVSEKILSALRIPDVGNLISQVVMKSKNDQGTNKKCNLONNNGNDIGFIFGHOFNNIAK 1260
DB 1201 GVSEKILSALRIPDVGNLISQVVMKSKNDQGTNKKCNLONNNGNDIGFIFGHOFNNIAK 1260
QY 1261 LVASNNYNRQIERSRRTLGCSWEFIPVDDGWERPL 1296
DB 1261 LVASNNYNRQIERSRRTLGCSWEFIPVDDGWERPL 1296

RESULT 5

US-09-255-829-8
; Sequence 8, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESS: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893

241 RVFVNTNAYEMSGLEVSPEELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
241 RVFVNTNAYEMSGLEVSPEELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDKLFDKLYKMLTEIYEDNFKPKV 360
301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDKLFDKLYKMLTEIYEDNFKPKV 360
361 LNRKTYLNFKAQVFKINIVPKVNTIYDGNLRLNTLAAFNFGONTINNMFNFKLNFT 420
361 LNRKTYLNFKAQVFKINIVPKVNTIYDGNLRLNTLAAFNFGONTINNMFNFKLNFT 420
421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
481 ITSDTNEAAEENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERPFG 540
481 ITSDTNEAAEENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERPFG 540
541 KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
541 KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
661 AVILLEFIPETAIPLVGTGTFALVSIAKVLTVQIDNALSKRNEKWDDEVYKIVTNWLAK 720
661 AVILLEFIPETAIPLVGTGTFALVSIAKVLTVQIDNALSKRNEKWDDEVYKIVTNWLAK 720
721 VNTQIDILIRKMKKEALENOAEATKAIINYQVNYTEEBEKNININIDDLSSKLNESINKA 780
721 VNTQIDILIRKMKKEALENOAEATKAIINYQVNYTEEBEKNININIDDLSSKLNESINKA 780
781 MININKEFNQCSVSYLNMSPYGVKRLDQFASLKDALLKYIYDNRGTLIGQVDRLLKOK 840
781 MININKEFNQCSVSYLNMSPYGVKRLDQFASLKDALLKYIYDNRGTLIGQVDRLLKOK 840
841 VNNLTSTDIIPQLSKYVDNQRLSTFTTEYIK 871
841 VNNLTSTDIIPQLSKYVDNQRLSTFTTEYIK 871

RESULT 7

US-09-255-829-26
; Sequence 26, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273

FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 871 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-26
Query Match 66.1%; Score 4507; DB 4; Length 871;
Best Local Similarity 99.8%; Pred. No. 6.6e-295;
Matches 869; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MPFVNFQVNFYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTNPEEGDLN 60
DB 1 MQFVNFQVNFYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTNPEEGDLN 60
QY 61 PPEAKQVPVSYSDYSLTDNEKNYLGKVTGLFERIYSTDLGRMLLTISIVRGIPFWGG 120
DB 61 PPEAKQVPVSYSDYSLTDNEKNYLGKVTGLFERIYSTDLGRMLLTISIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNVIIGPSADIIOFECKSGFHEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQPDGSGYRSEELNVIIGPSADIIOFECKSGFHEVLNLTNGY 180
QY 181 GSTQYRFSDDPTFGFEESLEVDNPLLGAKGKATDPAVTLAHELIIYAGHRLYGIANPN 240
DB 181 GSTQYRFSDDPTFGFEESLEVDNPLLGAKGKATDPAVTLAHELIIYAGHRLYGIANPN 240
QY 241 RVFKVNTNAYEMSGLEVSPEELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSPEELTFGCHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
QY 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDKLFDKLYKMLTEIYEDNFKPKV 360
DB 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSFGKSVKDKLFDKLYKMLTEIYEDNFKPKV 360
QY 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGNLRLNTLAAFNFGONTINNMFNFKLNFT 420
DB 361 LNRKTYLNFKAQVFKINIVPKVNTIYDGNLRLNTLAAFNFGONTINNMFNFKLNFT 420
QY 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
DB 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGE 480
QY 481 ITSDTNEAAEENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERPFG 540
DB 481 ITSDTNEAAEENISLDLIQOYYLTFNFDNEPENISLENSSDIIGOLELMPNTERPFG 540
QY 541 KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KYELDKYTWPHYLRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
DB 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALLFSG 660
QY 661 AVILLEFIPETAIPLVGTGTFALVSIAKVLTVQIDNALSKRNEKWDDEVYKIVTNWLAK 720
DB 661 AVILLEFIPETAIPLVGTGTFALVSIAKVLTVQIDNALSKRNEKWDDEVYKIVTNWLAK 720
QY 721 VNTQIDILIRKMKKEALENOAEATKAIINYQVNYTEEBEKNININIDDLSSKLNESINKA 780

Db 721 VNTQIDILIRKMKKALENQAATKAIINYQYNTVEEKNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCVSYLWNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
Db 781 MININKFLNOCVSYLWNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQVDRLLKDK 840
Qy 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871
Db 841 VNNLTSTDIPFQLSKYVDNORLLSTFTTEYIK 871

RESULT 8
US-09-255-829-18
; Sequence 18, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1013 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-18

Query Match 66.0%; Score 4501; DB 4; Length 1013;
Best Local Similarity 95.5%; Pred. No. 2.1e-294;
Matches 877; Conservative 3; Mismatches 12; Indels 26; Gaps 2;

Qy 1 MPFVNKQPNYKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTNPEEGDLN 60
Db 1 MQFVNKQPNYKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIERDFTNPEEGDLN 60
Qy 61 PPPAKQVPVSYDYSTYLTSTNEDKDYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPAKQVPVSYDYSTYLTSTNEDKDYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIQFECKSPGHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPGSVRSBELNVIIGPSADIIQFECKSPGHEVLNLTNGY 180

Qy 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPDAVTLAHELIYVAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPDAVTLAHELIYVAGHRLYGIAPNP 240
Qy 241 RVKQVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKKDIATSLNKA 300
Db 241 RVKQVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKKDIATSLNKA 300
Qy 301 KSIIVGTTASLOYKMNVPFKEKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYYEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYKMNVPFKEKYLLEDSTSGKFSVDKLFKDFKLYKMLTEIYYEDNFVKPFKV 360
Qy 361 LNRKTYLNFDPKAVPKINIVPKVNYTYDGFNLNLTNLAANFNGQNTINNNMFTKLKXFT 420
Db 361 LNRKTYLNFDPKAVPKINIVPKVNYTYDGFNLNLTNLAANFNGQNTINNNMFTKLKXFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSKLDKGNK-----ALNDLCIKVNNWDLPFSPSEDNFTN 473
Db 421 GLPEFYKLLCVRGIIITSKTSKLDKGNKIEGRCDGALNDLCIKVNNWDLPFSPSEDNFTN 480
Qy 474 DLNKGBEITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISLENLSSDIIGOLELMPN 533
Db 481 DLNKGBEITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISLENLSSDIIGOLELMPN 540
Qy 534 IERPPNGKYELDKYTMFHYLRAQEPFEGKSRIALTNSVNEALLNPSRVYTFPSSDYVKK 593
Db 541 IERPPNGKYELDKYTMFHYLRAQEPFEGKSRIALTNSVNEALLNPSRVYTFPSSDYVKK 600
Qy 594 VNKA TEAAMELGWVEQLVYDFTDETSEVSTTDKTDITIIIPYIGPALNIGNMLYKDDFV 653
Db 601 VNKA TEAAMELGWVEQLVYDFTDETSEVSTTDKTDITIIIPYIGPALNIGNMLYKDDFV 660
Qy 654 GALIFSGAVILLEFIPFIAIPVLGTFAVSYIAKVKLTQVTDNALSKRNEKWDVYKII 713
Db 661 GALIFSGAVILLEFIPFIAIPVLGTFAVSYIAKVKLTQVTDNALSKRNEKWDVYKII 720
Qy 714 VTNWLAKVNTQIDILIRKMKKALENQAATKAIINYQYNTVEEKNINFNIDDLSSKL 773
Db 721 VTNWLAKVNTQIDILIRKMKKALENQAATKAIINYQYNTVEEKNINFNIDDLSSKL 780
Qy 774 NESINKAMININKFLNOCVSYLWNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQ 833
Db 781 NESINKAMININKFLNOCVSYLWNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLLIGQ 840
Qy 834 VDRLLKDKVNTLSTDIPFQLSKYVDNORLLSTFTTEYIK-----NII 874
Db 841 VDRLLKDKVNTLSTDIPFQLSKYVDNORLLSTFTTEYIKSGLNSPGAAHYAQHDEAVDNKF 900

875 NTSILNURYESNHLIDLS 892
901 NKEQONAFYEILHLPNLN 918

RESULT 9
US-09-255-829-6
; Sequence 6, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-6

Query Match 65.9%; Score 4495; DB 4; Length 873;
Best Local Similarity 99.3%; Pred. No. 4.2e-294;
Matches 867; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPFVKQFNKYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Db 3 MEFVKNQFNKYKDPVNGVDIAIYIKIPKYGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 62
Qy 61 PPPEAKQVPVSYDYSTVLSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
Db 63 PPPEAKQVPVSYDYSTVLSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 122
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIFQCKSFSGHEVLNLTNGY 180
Db 123 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIFQCKSFSGHEVLNLTNGY 182
Qy 181 GSTQYIRFSDFTFGPESLEVDNPLLGAGKFTDPAVTLAHLIYAGHRLYGIAPNP 240
Db 183 GSTQYIRFSDFTFGPESLEVDNPLLGAGKFTDPAVTLAHLIYAGHRLYGIAPNP 242
Qy 241 RVKVNNTNAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNNKFDIATSLNKA 300
Db 243 RVKVNNTNAYEMSGLEVSFEELTFGCHDAKFTDSIQENEFRLYYNNKFDIATSLNKA 302
Qy 301 KSIIVGTTASLQYMKNVFKEKYLSEDTSGKFSVDKLFKDLKLYKMLTBIYTEDNFVKPFKV 360
Db 303 KSIIVGTTASLQYMKNVFKEKYLSEDTSGKFSVDKLFKDLKLYKMLTBIYTEDNFVKPFKV 362
Qy 361 LNRKTYLNFKAVPKINIVPKVNTIYDGNFLRNLTNLAANFNGONTINNNNFTKLKQFT 420
Db 363 LNRKTYLNFKAVPKINIVPKVNTIYDGNFLRNLTNLAANFNGONTINNNNFTKLKQFT 422
Qy 421 GLPEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKYNNNDLFPSPEDNFTNDLNKGE 480
Db 423 GLPEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKYNNNDLFPSPEDNFTNDLNKGE 482
Qy 481 ITSNTNIEAAEENISLDLIQQYLYTFNFDNENPENISNLSDDIIGQELMNPNIERPNG 540
Db 483 ITSNTNIEAAEENISLDLIQQYLYTFNFDNENPENISNLSDDIIGQELMNPNIERPNG 542
Qy 541 KKYELDKYTFHYLRAQEFBHGKSRIALTNSVNEALLNPGRVYTFPSSDYVKKVKNKATEA 600
Db 543 KKYELDKYTFHYLRAQEFBHGKSRIALTNSVNEALLNPGRVYTFPSSDYVKKVKNKATEA 602
Qy 601 AMFLGWEOQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 660
Db 603 AMFLGWEOQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGALIFSG 662

661 AVILLEPIETAIPLVLTGFALVSVYIANKVLTQVIDNALSKRNEKWDVEVKYIVTNMLAK 720
663 AVILLEPIETAIPLVLTGFALVSVYIANKVLTQVIDNALSKRNEKWDVEVKYIVTNMLAK 722
721 VNTQIDILIRKKMEALENOAEATKAIINYQVNOYVTEEEKNNINENIDDLSSKLNESINKA 780
723 VNTQIDILIRKKMEALENOAEATKAIINYQVNOYVTEEEKNNINENIDDLSSKLNESINKA 782
781 MININKELNOCVSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 840
783 MININKELNOCVSVYLMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTLIGQVDRCLKDK 842
841 VNNLTSTDIPFQLSKYVDNQRLLSSTFTTEYIK 871
843 VNNLTSTDIPFQLSKYVDNQRLLSSTFTTEYIK 873

RESULT 10
US-09-255-829-10
; Sequence 10, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 875 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-255-829-10

Query Match 65.9%; Score 4495; DB 4; Length 875;
Best Local Similarity 99.3%; Pred. No. 4.2e-294;
Matches 869; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 1 MPFVKQFNKYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Db 1 MQFVKQFNKYKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIRDTFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDYSTVLSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTVIRGIPFWGG 120
```


STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/255,829
FILING DATE: 23-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273
FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 953 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-14

Query Match 65.9%; Score 4494.5; DB 4; Length 953;
Best Local Similarity 98.0%; Pred. No. 5.2e-294;
Matches 869; Conservative 2; Mismatches 1; Indels 7; Gaps 1;

Qy 1 MPFVNKQPNYKDPVNGVDIAIKIPNAGQMOPVKAFFKHNKIWIPIRDTPTNPEEGDLN 60
Db 1 MQFVNKQPNYKDPVNGVDIAIKIPNAGQMOPVKAFFKHNKIWIPIRDTPTNPEEGDLN 60

Qy 61 PPEAKQVPVSYDSTYLTSTNEXNDYKGVTKLPERIYSTDLGRMLTSTVRCIPFWGG 120
Db 61 PPEAKQVPVSYDSTYLTSTNEXNDYKGVTKLPERIYSTDLGRMLTSTVRCIPFWGG 120

Qy 121 STIDTELKVIDTNCINVIQDGSYRSELNLVIGPSADIIQFCKSGFGEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQDGSYRSELNLVIGPSADIIQFCKSGFGEVLNLTNGY 180

Qy 181 GSTQYIRPSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240
Db 181 GSTQYIRPSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIYAGHRLYGIAPNP 240

Qy 241 RVFKVNTNAYEMSGLEVSFEELATFGHDAKFDISLQENEFRLYYNKKDIASTLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELATFGHDAKFDISLQENEFRLYYNKKDIASTLNKA 300

Qy 301 KSIIVGTTASLOYMNKVPKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYMNKVPKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYEDNFVKPFKV 360

Qy 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFGQNTENNNNFYKLNFT 420
Db 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFGQNTENNNNFYKLNFT 420

Qy 421 GLFSEYKLLCVRGIIITSKTSLDGYNK-----ALNDLCIKYNNNDLFPSPSEDFTN 473
Db 421 GLFSEYKLLCVRGIIITSKTSLDGYNKIGRCDCGALNDLCIKYNNNDLFPSPSEDFTN 480

Qy 474 DLNKGEEITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISLENLSSDIIGOLEMPN 533
Db 481 DLNKGEEITSDTNEAAEENISLDLIQOYLYLTFFNDEPENISLENLSSDIIGOLEMPN 540

Qy 534 IERFPNGKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 593

Db 541 IERFPNGKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKK 600

Qy 594 VNKATEAAMFLGWVEQLVYDFTDSTSEVSTTDDKIADITIIPIYGPALNIGNMLYKDDFV 653
Db 601 VNKATEAAMFLGWVEQLVYDFTDSTSEVSTTDDKIADITIIPIYGPALNIGNMLYKDDFV 660

Qy 654 GALIFSGAVILLEPIEIPALVLTGTFALSVYANKVLTVQTDIDNALSKEKMKWDEVYKI 713
Db 661 GALIFSGAVILLEPIEIPALVLTGTFALSVYANKVLTVQTDIDNALSKEKMKWDEVYKI 720

Qy 714 VTNWLAKVNTQIDILIRKMKKEALENOABATKAIINYYNQYTEEEKKNINFNIDDLSSKL 773
Db 721 VTNWLAKVNTQIDILIRKMKKEALENOABATKAIINYYNQYTEEEKKNINFNIDDLSSKL 780

Qy 774 NESINKAMININFLNOCVSYLMNSMIPYGVKRLDFDASLKDALDKYIYDNRGTLIGQ 833
Db 781 NESINKAMININFLNOCVSYLMNSMIPYGVKRLDFDASLKDALDKYIYDNRGTLIGQ 840

Qy 834 VDLRKQKVNNTLSTDIPIFQLSKYVDNQRLSTFTFYIKN 872
Db 841 VDLRKQKVNNTLSTDIPIFQLSKYVDNQRLSTFTFYIKS 879

RESULT 14
US-09-255-829-12
Sequence 12, Application US/09255829
Patent No. 6461617
GENERAL INFORMATION:
APPLICANT: Shone, Clifford Charles
APPLICANT: Quinn, Conrad Padraig
APPLICANT: Foster, Keith Alan
TITLE OF INVENTION: Recombinant Toxin Fragments
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/255,829
FILING DATE: 23-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02273
FILING DATE: 22-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,893
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-12

Query Match 65.9%; Score 4493.5; DB 4; Length 878;
Best Local Similarity 99.0%; Pred. No. 5.4e-294;
Matches 869; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

```
Qy 1 MPFVNQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
Db 1 MQFVNQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGSRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGSRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIATSLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIATSLNKA 300
Qy 301 KSIIVGTTASIQYMKNVFKEKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASIQYMKNVFKEKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFT 420
Db 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK-----ALNDLCIKVNNWDLFFSPSEDNFTN 473
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKIEGRCDGALNDLCIKVNNWDLFFSPSEDNFTN 480
Qy 474 DLNKGEEITSDTNEAAEENISLDLIQOYVLTFTNFDNEPENISTENLSSDIIIGOLELMPN 533
Db 474 DLNKGEEITSDTNEAAEENISLDLIQOYVLTFTNFDNEPENISTENLSSDIIIGOLELMPN 540
Qy 534 IERFPNGKVELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKK 593
Db 534 IERFPNGKVELDKYTMFHYLRAQEFEGHKSRIALTNSVNEALLNPSRVYTFSSDYVKK 600
Qy 594 VNKAATEAMFLGWVEQLVYDFTDSTSEVSTDKIADITIIPIYGPALNIGNMLYKODFV 653
Db 594 VNKAATEAMFLGWVEQLVYDFTDSTSEVSTDKIADITIIPIYGPALNIGNMLYKODFV 660
Qy 654 GALIFSGAVILLEFIPETAIPLVLTFAVSVIANKVLTVQTDNALSKRNEKWDVYKII 713
Db 654 GALIFSGAVILLEFIPETAIPLVLTFAVSVIANKVLTVQTDNALSKRNEKWDVYKII 720
Qy 714 VTNWLAKVNTQIDILIRKMKKEALENQAETKAIINYOYQVNYTEBEKKNINFNIDDLSSKL 773
Db 714 VTNWLAKVNTQIDILIRKMKKEALENQAETKAIINYOYQVNYTEBEKKNINFNIDDLSSKL 780
Qy 774 NESINKAMINIKFNLQCSVSYLWMSMIPYGVKREDFDASLKDALLKYIYDNRGTLLIG 833
Db 774 NESINKAMINIKFNLQCSVSYLWMSMIPYGVKREDFDASLKDALLKYIYDNRGTLLIG 840
Qy 834 VDLRKDKVNTLSTDDIPQLSKYVDNORLLSTFTTEYIK 871
Db 841 VDLRKDKVNTLSTDDIPQLSKYVDNORLLSTFTTEYIK 878
```

RESULT 15

US-09-288-326A-7
; Sequence 7, Application US/09288326A
; Patent No. 6776990
; GENERAL INFORMATION:
; APPLICANT: Steward, Lance E.
; APPLICANT: Aoki, K. Roger
; APPLICANT: Sachs, George
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Pancreatitis
; FILE REFERENCE: 17282

; CURRENT APPLICATION NUMBER: US/09/288,326A
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-288-326A-7

Query Match 34.4%; Score 2348; DB 4; Length 448;
Best Local Similarity 99.8%; Pred. No. 5.2e-150; Indels 0; Gaps 0;
Matches 447; Conservative 1; Mismatches 0

```
Qy 1 MPFVNQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
Db 1 MPFVNQFNKYKDPVNGVDIAYIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGSRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGSRSEELNLIIGPSADIIOPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGFEESELDVNTPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIATSLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIATSLNKA 300
Qy 301 KSIIVGTTASIQYMKNVFKEKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASIQYMKNVFKEKYLSEDTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFT 420
Db 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGFNLRLNTNLAANFNGQNTENNMFNFKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK 448
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK 448
```

Search completed: January 31, 2005, 14:09:56
Job time : 48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:04 ; Search time 174 Seconds
(without alignments)
2703.441 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKDPVNGVDIA.....EFIPVDGNGERPLHHHHH 1302

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6804	99.8	1306	15	US-10-205-516-16
2	6768	99.2	1296	15	US-10-452-024-160
3	6768	99.2	1296	15	US-10-205-516-2
4	6764	99.2	1296	15	US-10-452-024-159
5	6764	99.2	1296	17	US-10-821-669-1
6	6759	99.1	1295	15	US-09-726-949A-1
7	6759	99.1	1295	15	US-10-452-024-157
8	6756	99.0	1296	14	US-10-272-898-28
9	6756	99.0	1296	14	US-10-354-774-28
10	6756	99.0	1296	14	US-10-371-012-28
11	6756	99.0	1296	15	US-10-452-024-158
12	6756	99.0	1296	16	US-10-729-122-28
13	6756	99.0	1296	16	US-10-729-039-28

14	6756	99.0	1296	17	US-10-729-527-28	Sequence 28, Appl
15	6756	99.0	1296	17	US-10-727-898-28	Sequence 28, Appl
16	6756	99.0	1296	17	US-10-728-696-28	Sequence 28, Appl
17	6158	90.3	1296	15	US-10-452-024-1	Sequence 1, Appl
18	6158	90.3	1296	15	US-10-452-024-130	Sequence 130, Appl
19	4515	66.2	871	9	US-09-255-829-8	Sequence 8, Appl
20	4515	66.2	871	14	US-10-241-596-8	Sequence 8, Appl
21	4512	66.1	871	14	US-10-241-596-155	Sequence 155, App
22	4512	66.1	873	14	US-10-241-596-163	Sequence 163, App
23	4512	66.1	873	14	US-10-241-596-165	Sequence 165, App
24	4509	66.1	873	14	US-10-241-596-169	Sequence 169, App
25	4507	66.1	871	9	US-09-255-829-2	Sequence 2, Appl
26	4507	66.1	871	9	US-09-255-829-26	Sequence 26, Appl
27	4507	66.1	871	14	US-10-241-596-2	Sequence 2, Appl
28	4507	66.1	871	14	US-10-241-596-26	Sequence 26, Appl
29	4507	66.1	871	14	US-10-241-596-153	Sequence 153, App
30	4507	66.1	871	15	US-10-452-024-154	Sequence 154, App
31	4507	66.1	873	14	US-10-241-596-149	Sequence 149, App
32	4507	66.1	873	14	US-10-241-596-151	Sequence 151, App
33	4507	66.1	879	14	US-10-241-596-159	Sequence 159, App
34	4507	66.1	1127	14	US-10-241-596-40	Sequence 40, Appl
35	4507	66.1	1129	14	US-10-241-596-42	Sequence 42, Appl
36	4504	66.0	873	14	US-10-241-596-167	Sequence 167, App
37	4501	66.0	1013	9	US-09-255-829-18	Sequence 18, Appl
38	4501	66.0	1013	14	US-10-241-596-18	Sequence 18, Appl
39	4495.5	65.9	1130	14	US-10-241-596-44	Sequence 44, Appl
40	4495.5	65.9	1130	14	US-10-241-596-139	Sequence 139, App
41	4495.5	65.9	1132	14	US-10-241-596-46	Sequence 46, Appl
42	4495	65.9	873	9	US-09-255-829-6	Sequence 6, Appl
43	4495	65.9	873	14	US-10-241-596-6	Sequence 6, Appl
44	4495	65.9	875	9	US-09-255-829-10	Sequence 10, Appl
45	4495	65.9	875	14	US-10-241-596-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-10-205-516-16
; Sequence 16, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; FILE REFERENCE: jzbtx1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1306
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-16

Query Match	99.8%;	Score 6804;	DB 15;	Length 1306;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1301;	Conservative 1;	Mismatches 0;	Indels 4;	Gaps 1;
Qy	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGQMPKAFKHNKIWIPIERDFTTPEGDNLN	60	
Db	1	MPFVNQFNKDPVNGVDIAIYIKIPNAGQMPKAFKHNKIWIPIERDFTTPEGDNLN	60	
Qy	61	PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLFEIYSTDIGRMLLTSGIVGIPWGG	120	
Db	61	PPPEAKQVPVSYDSTYSLTDNEKDNLYKGVTKLFEIYSTDIGRMLLTSGIVGIPWGG	120	
Qy	121	STDTTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOQECKSFGEHVLNLTNGY	180	
Db	121	STDTTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOQECKSFGEHVLNLTNGY	180	
Qy	181	GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKATDPAVTLAHLIYAGHRLYGIATNP	240	

Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAINPN 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOYMKNVFKEKYLSEDTSKGFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYMKNVFKEKYLSEDTSKGFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAKAVKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDAKAVKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNK---ALNDLCIKVNNWDLFPSPSEDNFTNDLN 476
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKIEGRANDLCIKVNNWDLFPSPSEDNFTNDLN 480
Qy 477 KGEIITSDTNEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 536
Db 481 KGEIITSDTNEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 540
Qy 537 FPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKN 596
Db 541 FPNKKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKN 600
Qy 597 ATEAAMFLGWVEQLVYDFTDTSVSTTKADIITIIPIYGALNIGNMLYKDDFVGAL 656
Db 601 ATEAAMFLGWVEQLVYDFTDTSVSTTKADIITIIPIYGALNIGNMLYKDDFVGAL 660
Qy 657 IFSGAVILLESPIEIALPVLTGFALVSIAKVLTVOTIDNALSIRNEKEDVEYKIVTN 716
Db 661 IFSGAVILLESPIEIALPVLTGFALVSIAKVLTVOTIDNALSIRNEKEDVEYKIVTN 720
Qy 717 WLAKVNTQIDILIRKMKAEALNOEAATKAIINYQYNTYBEEKNNINFNIDDLSSKLNES 776
Db 721 WLAKVNTQIDILIRKMKAEALNOEAATKAIINYQYNTYBEEKNNINFNIDDLSSKLNES 780
Qy 777 INKAMININFLNCSVSYLMNSMIPYGVKLEDFDASLKDALIKYIDNRGTILGOVDR 836
Db 781 INKAMININFLNCSVSYLMNSMIPYGVKLEDFDASLKDALIKYIDNRGTILGOVDR 840
Qy 837 LKDKVNTLTSDIPFQLSKYVDNORLLSTFTFYIKNIINTSILNRYESNHLIDLSRYAS 896
Db 841 LKDKVNTLTSDIPFQLSKYVDNORLLSTFTFYIKNIINTSILNRYESNHLIDLSRYAS 900
Qy 897 KINIGSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVYNSMYENFSTFWIRIPKYFNSI 956
Db 901 KINIGSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVYNSMYENFSTFWIRIPKYFNSI 960
Qy 957 SLANNEYTIINCWENNSGWKSYLNGEIIWTLQDTQEIQRVVRKYQVYKINISDYINRWIF 1016
Db 961 SLANNEYTIINCWENNSGWKSYLNGEIIWTLQDTQEIQRVVRKYQVYKINISDYINRWIF 1020
Qy 1017 VTIITNNLNNSKIYINGRLIDQKPISNLGINHASNNIMFKDGCGRDTHRYIKYFNLF 1076
Db 1021 VTIITNNLNNSKIYINGRLIDQKPISNLGINHASNNIMFKDGCGRDTHRYIKYFNLF 1080
Qy 1077 KELNEKEIKOLYDNQNSGILKDPWGDYLVQDKPYMLNLYDPNKYVDVNVNIGRGYMYL 1136
Db 1081 KELNEKEIKOLYDNQNSGILKDPWGDYLVQDKPYMLNLYDPNKYVDVNVNIGRGYMYL 1140
Qy 1137 KGPGRSVWNTNIIYLNSSIIYGTKEPIIKKYASGNKDNVNRNDRVYINNVKKNERYLATN 1196
Db 1141 KGPGRSVWNTNIIYLNSSIIYGTKEPIIKKYASGNKDNVNRNDRVYINNVKKNERYLATN 1200
Qy 1197 ASQAGVEKILSALBIPDVGNLSQVVVMKSNKQDQITNCKMNLQDNGNDIGTGFHQFN 1256
Db 1201 ASQAGVEKILSALBIPDVGNLSQVVVMKSNKQDQITNCKMNLQDNGNDIGTGFHQFN 1260
Qy 1257 NIAKLVASNWNQRQIERSRSLTGCSEWEIFPVDDGGERPLHHHHH 1302

Db 1261 NIAKLVASNWNQRQIERSRSLTGCSEWEIFPVDDGGERPLHHHHH 1306

RESULT 2

US-10-452-024-160
; Sequence 160, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transepithelial Molecular Transport
; FILE REFERENCE: 9855-96V1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 160
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; US-10-452-024-160

Query Match 99.2%; Score 6768; DB 15; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGQMPVKAFKIHNKIWIIPERDFTTPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGQMPVKAFKIHNKIWIIPERDFTTPEEGDLN 60
Qy 61 PPEAKQPVSYVSDTSLSTDNKDNVKGVTKLFBRIYSTDLGRMLTISIVRGIPWGG 120
Db 61 PPEAKQPVSYVSDTSLSTDNKDNVKGVTKLFBRIYSTDLGRMLTISIVRGIPWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLIIGPSADIIOFECKSFGEHLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAINPN 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHLIHAHRLYGIAINPN 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFTDSLOENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOYMKNVFKEKYLSEDTSKGFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYMKNVFKEKYLSEDTSKGFSVDKLFKDFKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAKAVKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDAKAVKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 480
Db 421 GLPEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 480
Qy 481 ITSDTNIEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 540
Db 481 ITSDTNIEAAEENISLDLIQQYVLTFFNFDNEPENISLENSSDIIGOLELMPNIE 540
Qy 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA 600
Db 541 KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDTSVSTTKADIITIIPIYGALNIGNMLYKDDFVGALIFSG 660
Db 601 AMFLGWVEQLVYDFTDTSVSTTKADIITIIPIYGALNIGNMLYKDDFVGALIFSG 660

```

Qy 661 AVILLEPIEIAIPVLGTFALVSIAKVLTVQTIIDNALSKRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEPIEIAIPVLGTFALVSIAKVLTVQTIIDNALSKRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNTLSTDIIPFOLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VNTLSTDIIPFOLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNISILNN 960
Db 901 GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNISILNN 960
Qy 961 EYTIINCMMNSGKWSLNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTIINCMMNSGKWSLNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKOLDYDQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR 1140
Db 1081 EKEIKOLDYDQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR 1140
Qy 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNEKRYLATNASQA 1200
Db 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNEKRYLATNASQA 1200
Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGTGFQHFNNIAK 1260
Db 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKKNLQDNGNDIGTGFQHFNNIAK 1260
Qy 1261 LVASNNVNRQIERSSRTLGCSEWEPVDDGGERPL 1296
Db 1261 LVASNNVNRQIERSSRTLGCSEWEPVDDGGERPL 1296

```

RESULT 3

```

US-10-205-516-2
; Sequence 2, Application US/10205516
; Publication No. US20040018589A1
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; FILE REFERENCE: JZb2x1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1296
; TYPE: PR1
; ORGANISM: Clostridium botulinum
US-10-205-516-2

```

```

Query Match 99.2%; Score 6768; DB 15; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps :0;

```

```

Qy 1 MPFVNKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKHKNKIWIIPERDFTNPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKHKNKIWIIPERDFTNPEEGDLN 60
Qy 61 PPPEAKQVPVSYDSTYLDNEKDNLYLKGVTKLFRYIYSTDLGRMLLTSIVRGIPFWGG 120

```

```

Db 61 PPPEAKQVPVSYDSTYLDNEKDNLYLKGVTKLFRYIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLVLIIGPSADIIQPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNLVLIIGPSADIIQPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAVYEMSGLEVSPFELRTFGGHDAKFIDSLOENEFRLYYNKKPKDASTINKA 300
Db 241 RVFKVNTNAVYEMSGLEVSPFELRTFGGHDAKFIDSLOENEFRLYYNKKPKDASTINKA 300
Qy 301 KSVIGTTASIQYMKNVFKERYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSVIGTTASIQYMKNVFKERYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFDPKAVFKINIVPKVNYTYDGNPLNLTNLAANFNGQNTENNMTFKLNFT 420
Db 361 LNRKTYLNFDPKAVFKINIVPKVNYTYDGNPLNLTNLAANFNGQNTENNMTFKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKTSKLDKYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
Db 421 GLFEFYKLLCVRGIIITSKTSKLDKYNKALNDLCIKVNNWDLFPSPSEDNFTDNLKGE 480
Qy 481 ITSDDTNIIEAENISLDLIIQOYYLTFNFDNENPISNIENLSSDIIGOLELMPNIERPNG 540
Db 481 ITSDDTNIIEAENISLDLIIQOYYLTFNFDNENPISNIENLSSDIIGOLELMPNIERPNG 540
Qy 541 KXYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Db 541 KXYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKODFVGLIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKODFVGLIFSG 660
Qy 661 AVILLEPIEIAIPVLGTFALVSIAKVLTVQTIIDNALSKRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEPIEIAIPVLGTFALVSIAKVLTVQTIIDNALSKRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYQVQYVTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Db 781 MININKFLNOCSSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTLIGQVDRLLKDK 840
Qy 841 VNTLSTDIIPFOLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VNTLSTDIIPFOLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNISILNN 960
Db 901 GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVNSMYENFSTFWIRIPKYFNISILNN 960
Qy 961 EYTIINCMMNSGKWSLNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTIINCMMNSGKWSLNYGEIITWLTQDTEIKQKVVFYKYSQMINISDYINRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKOLDYDQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR 1140
Db 1081 EKEIKOLDYDQNSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNNVGVIRGYMYLKGPR 1140
Qy 1141 GSVMTNIIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNEKRYLATNASQA 1200

```

Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALETPDVGNSQVVMKSKNDQGTNNKKNLQDNGNDIGFIFGHOFNNIAK 1260
Db 1201 GVEKILSALETPDVGNSQVVMKSKNDQGTNNKKNLQDNGNDIGFIFGHOFNNIAK 1260
Qy 1261 LVASWYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296
Db 1261 LVASWYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296

RESULT 4
US-10-452-024-159
; Sequence 159, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Trans epithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 159
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-452-024-159

Query Match 99.2%; Score 6764; DB 15; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1294; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDFTNPEEGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAIKIPNAGOMQPVKAFKIHNNKIWIPIERDFTNPEEGDLN 60
Qy 61 PPPEAKQPVSYDSTYLTSTNEKDNVKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPPEAKQPVSYDSTYLTSTNEKDNVKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPGSYRSELNLVITGPSADIIQFCKSPGHEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPGSYRSELNLVITGPSADIIQFCKSPGHEVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIIYAGHRLYGIAINPN 240
Db 181 GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIIYAGHRLYGIAINPN 240
Qy 241 RVFKVNNAYEMSGLEVPSELRTFGGHDAKFIDSLQENEFRLYYNNKFKDIASTLNKA 300
Db 241 RVFKVNNAYEMSGLEVPSELRTFGGHDAKFIDSLQENEFRLYYNNKFKDIASTLNKA 300
Qy 301 KSIQVGTASLOYKMNVPKEKYLLEDTSKGFSVDKLKFDKLYKMLTEIYTEDNFVKFKV 360
Db 301 KSIQVGTASLOYKMNVPKEKYLLEDTSKGFSVDKLKFDKLYKMLTEIYTEDNFVKFKV 360
Qy 361 LNRKTYLNFDKAVKINIVPKVNTIYDGFNLRLNLAANFGQNTENNMMFTKLNFT 420
Db 361 LNRKTYLNFDKAVKINIVPKVNTIYDGFNLRLNLAANFGQNTENNMMFTKLNFT 420
Qy 421 GLFFBYKLLCVRGITTSKSLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGE 480
Db 421 GLFFBYKLLCVRGITTSKSLDKGYNKALNDLCIKVNNWDLFPSPSDNFTNDLNKGE 480
Qy 481 ITSDNTIEAABENISLDLIQQYLTFTNFDNPNENISLNLSSDIIGOLELMPNTERPFG 540
Db 481 ITSDNTIEAABENISLDLIQQYLTFTNFDNPNENISLNLSSDIIGOLELMPNTERPFG 540

Qy 541 KKVELDKYTMFHYLRAQEFPHGKSRIALTNSVNEALNPSRVYTFPSSDVYKKVKNKATEA 600
Db 541 KKVELDKYTMFHYLRAQEFPHGKSRIALTNSVNEALNPSRVYTFPSSDVYKKVKNKATEA 600
Qy 601 AMFLGWEOQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG 660
Db 601 AMFLGWEOQVYDFTDETSEVSTTDKIADITIIPIYGPALNIGNMLYKDDFVGCALIFSG 660
Qy 661 AVILLBEPIEAIAPVLGTFFALVSYIANKVLTVQIDNALSKNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLBEPIEAIAPVLGTFFALVSYIANKVLTVQIDNALSKNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKKMEALENQAEATKAIINYQVQYTEBEKNININPIDDLSKLNESINKA 780
Db 721 VNTQIDILIRKKMEALENQAEATKAIINYQVQYTEBEKNININPIDDLSKLNESINKA 780
Qy 781 MININKELNOCSSYLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTIGQVDRUKDK 840
Db 781 MININKELNOCSSYLYMNSMIPYGVKLEDFDASLKDALLKYIYDNRGTIGQVDRUKDK 840
Qy 841 VNNLTSTDIQFQSKYVDNORLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Db 841 VNNLTSTDIQFQSKYVDNORLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKNOIQLPNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNISILNN 960
Db 901 GSKVNFDPIDKNOIQLPNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNISILNN 960
Qy 961 EYTIINCWENNSGWKVSILNYGEIITWLTQDTEIKQRVVFKYSOMINISDYNRWIFVTIT 1020
Db 961 EYTIINCWENNSGWKVSILNYGEIITWLTQDTEIKQRVVFKYSOMINISDYNRWIFVTIT 1020
Qy 1021 NNRLNNSKIYINGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWIYFNLPDKELN 1080
Db 1021 NNRLNNSKIYINGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWIYFNLPDKELN 1080
Qy 1081 EKEIKOLYDNOSNGILKDFWGYLDYQDKPYMNLNLDPNKYVDVNVNVRGIRGYMYLKGPR 1140
Db 1081 EKEIKOLYDNOSNGILKDFWGYLDYQDKPYMNLNLDPNKYVDVNVNVRGIRGYMYLKGPR 1140
Qy 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALETPDVGNSQVVMKSKNDQGTNNKKNLQDNGNDIGFIFGHOFNNIAK 1260
Db 1201 GVEKILSALETPDVGNSQVVMKSKNDQGTNNKKNLQDNGNDIGFIFGHOFNNIAK 1260
Qy 1261 LVASWYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296
Db 1261 LVASWYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296

RESULT 5
US-10-821-669-1
; Sequence 1, Application US/10821669
; Publication No. US20040265935A1
; GENERAL INFORMATION:
; APPLICANT: Atassi, M. Zouhair
; TITLE OF INVENTION: Botulinum Toxin A Peptides and Methods
; TITLE OF INVENTION: of Predicting and Reducing Immunoresistance to Botulinum
; FILE REFERENCE: 66872-039
; CURRENT APPLICATION NUMBER: US/10/821,669
; CURRENT FILING DATE: 2004-04-09
; PRIOR APPLICATION NUMBER: US 60/462,754
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum

US-10-821-669-1

Query Match 99.2%; Score 6764; DB 17; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1294; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPFVKQPNYKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDFTNPEEGDLN 60
Db 1 MPFVKQPNYKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDYSTYLSSTNEKDNLYKGYTLKFERIYSTDLGRMLLTSIVRGIPFWGG 120
Db 61 PPEAKQVPVSYDYSTYLSSTNEKDNLYKGYTLKFERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIATNP 240
Db 181 GSTQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIATNP 240
Qy 241 RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDLKYLKMLTEIYEDNFVKPFKV 360
Db 301 KSIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDLKYLKMLTEIYEDNFVKPFKV 360
Qy 361 LNRKTYLNFDAVKFKNIPKVNVTIYDGNLRTNLAANPONGQNTINNMFVKLNFT 420
Db 361 LNRKTYLNFDAVKFKNIPKVNVTIYDGNLRTNLAANPONGQNTINNMFVKLNFT 420
Qy 421 GLPEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKVNNWDLFPSPEDNFTDNLNGEE 480
Db 421 GLPEFYKLLCVRGIIITKTSKLDGYNKALNDLCIKVNNWDLFPSPEDNFTDNLNGEE 480
Qy 481 ITSDTNEAEBENISLDLIQOYLYTFNFDNENPENSIEIENSSDIIGOLELMPNTERP 540
Db 481 ITSDTNEAEBENISLDLIQOYLYTFNFDNENPENSIEIENSSDIIGOLELMPNTERP 540
Qy 541 KKYELDKYTHPYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFPSSDYVKKYNKATEA 600
Db 541 KKYELDKYTHPYLRAQEFEGKGRIALTNSVNEALLNPSRVYTFPSSDYVKKYNKATEA 600
Qy 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKDDFVGCALIFSG 660
Db 601 AMFLGWEOQLVYDFTDETSEVSTTDKADIITIIPIYGPALNIGMLYKDDFVGCALIFSG 660
Qy 661 AVILLEFIPEITAIPIVLGTFFALVSIAKVLTVQITIDNALSKEKDEVKYIIVTNLAK 720
Db 661 AVILLEFIPEITAIPIVLGTFFALVSIAKVLTVQITIDNALSKEKDEVKYIIVTNLAK 720
Qy 721 VNTQIDILIRKWKALENQABATKAIINYQVQYTEBEKKNININIDLSKLNESINKA 780
Db 721 VNTQIDILIRKWKALENQABATKAIINYQVQYTEBEKKNININIDLSKLNESINKA 780
Qy 781 MININKFLNOCVSYLWNSMIPYGVKLEDDASIKOALLKYIYDNRGTLIGQVDRUKDK 840
Db 781 MININKFLNOCVSYLWNSMIPYGVKLEDDASIKOALLKYIYDNRGTLIGQVDRUKDK 840
Qy 841 VVNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNRYESNHLIDLRSYASKINI 900
Db 841 VVNTLSTDIPQLSKYVDNQRLSTFTTEYIKNIINTSILNRYESNHLIDLRSYASKINI 900
Qy 901 GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTLINCMENNSGKWSLNYGEIITWLODQTEIKORVVVKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTLINCMENNSGKWSLNYGEIITWLODQTEIKORVVVKYSQMINISDYINRWIFVTIT 1020

Qy 1021 NNRLLNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDEKLN 1080
Db 1021 NNRLLNSKIYINGRLIDQKPISNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDEKLN 1080
Qy 1081 EKEIKOLYDNQNSGILKDFWGDYQYDKPYMNLNDPNKYVDVNNVGIRGYMYLKGPR 1140
Db 1081 EKEIKOLYDNQNSGILKDFWGDYQYDKPYMNLNDPNKYVDVNNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTTNIYLNSSLYRGTFKFIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNSSLYRGTFKFIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNKCRLNLODNGNDIGFHFQFNNAK 1260
Db 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNKCRLNLODNGNDIGFHFQFNNAK 1260
Qy 1261 LVASNNYNRQIERSSRTLGCSEWFIIPVDDGWERPL 1296
Db 1261 LVASNNYNRQIERSSRTLGCSEWFIIPVDDGWERPL 1296

RESULT 6

US-09-726-949A-1
; Sequence 1, Application US/09726949A
; Patent No. US20020137886A1
; GENERAL INFORMATION:
; APPLICANT: Allergan, Inc.
; APPLICANT: Lin, Wei-Jen
; APPLICANT: Aoki, Kei Roger
; APPLICANT: Steward, Lance E.
; TITLE OF INVENTION: NEUROTOXINS WITH ENHANCED TARGET
; TITLE OF INVENTION: SPECIFICITY
; FILE REFERENCE: 36121-20002.00
; CURRENT APPLICATION NUMBER: US/09/726,949A
; CURRENT FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1295
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-726-949A-1

Query Match 99.1%; Score 6759; DB 9; Length 1295;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PFVVKQPNYKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDFTNPEEGDLN 61
Db 1 PFVVKQPNYKDPVNGVDIAVIKIPNAGOMQPVKAFKHNNKIWIIPERDFTNPEEGDLN 60
Qy 62 PPEAKQVPVSYDYSTYLSSTNEKDNLYKGYTLKFERIYSTDLGRMLLTSIVRGIPFWGG 121
Db 61 PPEAKQVPVSYDYSTYLSSTNEKDNLYKGYTLKFERIYSTDLGRMLLTSIVRGIPFWGG 120
Qy 122 TIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 181
Db 121 TIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIIQFECKSFGEHVLNLTNGY 180
Qy 182 STQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIATNP 241
Db 181 STQYIRFSPDFTFGFESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIATNP 240
Qy 242 VFQVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 301
Db 241 VFQVNTNAYYEMSGLEVSPEELRTFGHDAKFDISLQENEFRLYYNKKFDIASTLNKA 300
Qy 302 SIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDLKYLKMLTEIYEDNFVKPFKV 361
Db 301 SIIVGTTASLOYMKNVKFKYLLSEDTSGKFSVDKLFKDLKYLKMLTEIYEDNFVKPFKV 360
Qy 362 NRTYLNFDKAVKFINIVKVNVTIYDGNLRTNLAANPONGQNTINNMFVKLNFT 421

Db 361 NRKTYLNFDAVFKINIVPKVNYTIDYGFNLNTNLAANFNGQNTENNNTFTKLNFTG 420
Qy 422 LFEYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 481
Db 421 LFEYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 480
Qy 482 TSDTNIEAABENISLDLIQOYLTFFNPNPENISIEENLSSDIIGOLELMPNTERFPNGK 541
Db 481 TSDTNIEAABENISLDLIQOYLTFFNPNPENISIEENLSSDIIGOLELMPNTERFPNGK 540
Qy 542 KYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVNKATEAA 601
Db 541 KYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVNKATEAA 600
Qy 602 MFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDPVGALIFSGA 661
Db 601 MFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDPVGALIFSGA 660
Qy 662 VILLEFTEPEIAIPVLGTFALVSYIANKVLTQTIDNALSKNEKWDVYKIVTNWLAKV 721
Db 661 VILLEFTEPEIAIPVLGTFALVSYIANKVLTQTIDNALSKNEKWDVYKIVTNWLAKV 720
Qy 722 NTQIDLIRKMKKEALENOAEATKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 781
Db 721 NTQIDLIRKMKKEALENOAEATKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 780
Qy 782 ININFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGIGQVDRDKDV 841
Db 781 ININFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGIGQVDRDKDV 840
Qy 842 NNTLSTDPFQOLSKYVDNRQLSTFTYIKNIINTSILNLYESNHLIDLSRYASKINIG 901
Db 841 NNTLSTDPFQOLSKYVDNRQLSTFTYIKNIINTSILNLYESNHLIDLSRYASKINIG 900
Qy 902 SKVNFDPIDKNOIQLFNLESSKIEVILKNAIVNYSMYENFSTFWIRIPKYFNSISLNE 961
Db 901 SKVNFDPIDKNOIQLFNLESSKIEVILKNAIVNYSMYENFSTFWIRIPKYFNSISLNE 960
Qy 962 YTIINCMMNSGWKSLNYGIIWTLQDTQEI KORVVPKYSQMINISDIYNRWIPVITN 1021
Db 961 YTIINCMMNSGWKSLNYGIIWTLQDTQEI KORVVPKYSQMINISDIYNRWIPVITN 1020
Qy 1022 NRLANSKIYINGRLIDQKPI SNLGNHASNNIMFKLDCRDTYRVIWKYFNLFDEKELNE 1081
Db 1021 NRLANSKIYINGRLIDQKPI SNLGNHASNNIMFKLDCRDTYRVIWKYFNLFDEKELNE 1080
Qy 1082 KEIKDLYDNQNSGILKDFWGDYLOQYDKPYMLNLYDPNKVYDVNNVGIRGYMLKGRG 1141
Db 1081 KEIKDLYDNQNSGILKDFWGDYLOQYDKPYMLNLYDPNKVYDVNNVGIRGYMLKGRG 1140
Qy 1142 SVMTTNIYNSLSYRGTFPIIKKVASGNKDNIVRNDRVYINNVYKNEYRLATNASQAG 1201
Db 1141 SVMTTNIYNSLSYRGTFPIIKKVASGNKDNIVRNDRVYINNVYKNEYRLATNASQAG 1200
Qy 1202 VEKILSALIEIPDVGNLSQVVMKSKNDQGITNKCKMNLQDNNGNDIGFQHFQNNIAKL 1261
Db 1201 VEKILSALIEIPDVGNLSQVVMKSKNDQGITNKCKMNLQDNNGNDIGFQHFQNNIAKL 1260
Qy 1262 VASNNYNQIERSRSLTGCSEHFIIPVDDGWERPL 1296
Db 1261 VASNNYNQIERSRSLTGCSEHFIIPVDDGWERPL 1295

RESULT 7
US-10-452-024-157
; Sequence 157, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport

; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 157
; LENGTH: 1295
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; US-10-452-024-157

Query Match 99.1%; Score 6759; DB 15; Length 1295;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 2 PFYNKQPNYKDPNGVDIAYIKIPNAGOMQPVKAFKHNKIWIPIPRDPTNPEEGDLP 61
Db 1 PFYNKQPNYKDPNGVDIAYIKIPNAGOMQPVKAFKHNKIWIPIPRDPTNPEEGDLP 60
Qy 62 PPAKQVPVSYDYSTYLS TDNEKDNLYKGVTKLPERIYSTDLGRMLTTSIVRGIPFWGGS 121
Db 61 PPAKQVPVSYDYSTYLS TDNEKDNLYKGVTKLPERIYSTDLGRMLTTSIVRGIPFWGGS 120
Qy 122 TIDTELKVIDTNCINVIQPDGYSRSEBELNVLIIGPSADIIQPECKSPFGHEVLNLTNGYG 181
Db 121 TIDTELKVIDTNCINVIQPDGYSRSEBELNVLIIGPSADIIQPECKSPFGHEVLNLTNGYG 180
Qy 182 STQYIRSPDFTGFEESLVDNPLIGAGKFPATDPAVTLAHELIYAGHRLYGIATNPNR 241
Db 181 STQYIRSPDFTGFEESLVDNPLIGAGKFPATDPAVTLAHELIYAGHRLYGIATNPNR 240
Qy 242 VFKNVNAVYEMSGLEVSFEELRTFGHDAKFIDSLQENEFRLYYNNKFDIASTLNKAK 301
Db 241 VFKNVNAVYEMSGLEVSFEELRTFGHDAKFIDSLQENEFRLYYNNKFDIASTLNKAK 300
Qy 302 SIVGTTASLOYMKNVFKKYLKLLSEDTSGKFSVDKLPDKLYKMLTEIYTEDNFVKFPKVL 361
Db 301 SIVGTTASLOYMKNVFKKYLKLLSEDTSGKFSVDKLPDKLYKMLTEIYTEDNFVKFPKVL 360
Qy 362 NRKTYLNFDAVFKINIVPKVNYTIDYGFNLNTNLAANFNGQNTENNNTFTKLNFTG 421
Db 361 NRKTYLNFDAVFKINIVPKVNYTIDYGFNLNTNLAANFNGQNTENNNTFTKLNFTG 420
Qy 422 LFEYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 481
Db 421 LFEYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVANNWDLFFSPSDNFTNDLNKGEI 480
Qy 482 TSDTNIEAABENISLDLIQOYLTFFNPNPENISIEENLSSDIIGOLELMPNTERFPNGK 541
Db 481 TSDTNIEAABENISLDLIQOYLTFFNPNPENISIEENLSSDIIGOLELMPNTERFPNGK 540
Qy 542 KYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVNKATEAA 601
Db 541 KYELDKYTMFHYLRAQFEHKGSRIALTNSVNEALLNPSRVYTFSSDYVKVNKATEAA 600
Qy 602 MFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDPVGALIFSGA 661
Db 601 MFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKDDPVGALIFSGA 660
Qy 662 VILLEFTEPEIAIPVLGTFALVSYIANKVLTQTIDNALSKNEKWDVYKIVTNWLAKV 721
Db 661 VILLEFTEPEIAIPVLGTFALVSYIANKVLTQTIDNALSKNEKWDVYKIVTNWLAKV 720
Qy 722 NTQIDLIRKMKKEALENOAEATKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 781
Db 721 NTQIDLIRKMKKEALENOAEATKAIINYQYNYTEEEKNNINFNIDDLSSKLNESINKAM 780
Qy 782 ININFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGIGQVDRDKDV 841
Db 781 ININFLNQC SVSYLMNSMIPYGVKRLDFDASLKDALIKYIYDNRGTGIGQVDRDKDV 840

QY 842 NNTLSTDPFOLSKVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 901
DB NNTLSTDPFOLSKVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 900
QY 902 SKVNFDPIDKQIQOLFNLSSKIEVILKNAIVNYSMYENFSTSWIRIPKYNFNSISLNN 961
DB SKVNFDPIDKQIQOLFNLSSKIEVILKNAIVNYSMYENFSTSWIRIPKYNFNSISLNN 960
QY 962 YTIINCENNSGWKSVLNSYGBIIWTLQDTQBIKQVVPKYSQMINISDIYNRWIFVTIN 1021
DB YTIINCENNSGWKSVLNSYGBIIWTLQDTQBIKQVVPKYSQMINISDIYNRWIFVTIN 1020
QY 1022 NNLANSKIYINGRLIDQKPISNLGNIIHASNNIMFKLDCRDRTHRYIWKIFNLFDEKELNE 1081
DB NNLANSKIYINGRLIDQKPISNLGNIIHASNNIMFKLDCRDRTHRYIWKIFNLFDEKELNE 1080
QY 1082 KEIKDLDYDNQNSGILKDFWGDYLDQYDKPYMLNLYDPNKVYDVNNVGIKGYMYLKGPRG 1141
DB KEIKDLDYDNQNSGILKDFWGDYLDQYDKPYMLNLYDPNKVYDVNNVGIKGYMYLKGPRG 1140
QY 1142 SVMTTNIYLNSSLYRGTKFIKKYASGNKONIVRRNDRVYINVVVKNEYRLATNASQAG 1201
DB SVMTTNIYLNSSLYRGTKFIKKYASGNKONIVRRNDRVYINVVVKNEYRLATNASQAG 1200
QY 1202 VEKILSALEIPDVGNLSQVVMKSKNDGIIITNCKMNLQDNGNDIGFIGHQFNNTAKL 1261
DB VEKILSALEIPDVGNLSQVVMKSKNDGIIITNCKMNLQDNGNDIGFIGHQFNNTAKL 1260
QY 1262 VASWNYRQIRSRRTLGCSEWEIFPVDGWERPL 1296
DB VASWNYRQIRSRRTLGCSEWEIFPVDGWERPL 1295

RESULT 8

US-10-272-898-28
; Sequence 28, Application US/10272898
; Publication No. US20030162745A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/10/272,898
; FILING DATE: 04-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-272-898-28
Query Match 99.08; Score 6756; DB 14; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MPFVKQFNKYKDPVNGVDIAIYIKIPNAGQMPQKAFKHKNKIWIPIERDTFTNPEEGDLN 60
DB 1 MQFVKQFNKYKDPVNGVDIAIYIKIPNAGQMPQKAFKHKNKIWIPIERDTFTNPEEGDLN 60
QY 61 PPPEAKQPVSYSDYSTVLSTDNKKNYLKGVTKLFEIYISTDLGRMLLTSTVIRGIPWGG 120
DB 61 PPPEAKQPVSYSDYSTVLSTDNKKNYLKGVTKLFEIYISTDLGRMLLTSTVIRGIPWGG 120
QY 121 STIDTELKVIDTNCINVIQPGSVSRSEELNVIIGPSADIIQFECKSGFGEVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQPGSVSRSEELNVIIGPSADIIQFECKSGFGEVLNLTNGY 180
QY 181 GSTQYIRFSDPFTFGFEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAINPN 240
DB 181 GSTQYIRFSDPFTFGFEESLEVDTNPLLGAGKATDPAVTLAHELIVAGHRLYGIAINPN 240
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNPKFDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELRTFGCHDAKFDISLQENEFRLYYNPKFDIASTLNKA 300
QY 301 KSIIVGTTASLOYMKNVFKKYLLEDSTSGKFSVDKLFKDKLYKMLTETIYEDNFKVFKPV 360
DB 301 KSIIVGTTASLOYMKNVFKKYLLEDSTSGKFSVDKLFKDKLYKMLTETIYEDNFKVFKPV 360
QY 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGNFLRNTNLAANFNGQNTENNNTFKLNFT 420
DB 361 LNRKTYLNFDAVKFKNIVPKVNYTIYDGNFLRNTNLAANFNGQNTENNNTFKLNFT 420
QY 421 GLFEFYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
DB 421 GLFEFYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEE 480
QY 481 ITSDTNIEAAEENISLDLIQOYYLTFTNFDNEPENISLENSSDIIGOLELMPNTERPNG 540
DB 481 ITSDTNIEAAEENISLDLIQOYYLTFTNFDNEPENISLENSSDIIGOLELMPNTERPNG 540
QY 541 KYELDKYTMFHYLRAQEFHGHKSRIALTNSVNEALNPSRVYTFSSDYVKKYKATEA 600
DB 541 KYELDKYTMFHYLRAQEFHGHKSRIALTNSVNEALNPSRVYTFSSDYVKKYKATEA 600
QY 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
QY 661 AVILLEEPIEAIAPVLGTFAVSIANKVLTQVOTIDNALSKRNEKWDVYKIVTNWLAK 720
DB 661 AVILLEEPIEAIAPVLGTFAVSIANKVLTQVOTIDNALSKRNEKWDVYKIVTNWLAK 720
QY 721 VNTQIDILIRKKWEALENOAEATKAIINYOQVNYTEEEKNNINFINIDLSKLNESINKA 780
DB 721 VNTQIDILIRKKWEALENOAEATKAIINYOQVNYTEEEKNNINFINIDLSKLNESINKA 780
QY 781 MININKFLNQCYSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTIGQVDRLKDK 840
DB 781 MININKFLNQCYSVYLMNSMIPYGVKRELEDFDASLKDALLKYIYDNRGTIGQVDRLKDK 840
QY 841 VNTLSTDPFOLSKVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 900
DB 841 VNTLSTDPFOLSKVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLSRYASKINIG 900
QY 901 GSKVNFDPIDKQIQOLFNLSSKIEVILKNAIVNYSMYENFSTSWIRIPKYNFNSISLNN 960
DB 901 GSKVNFDPIDKQIQOLFNLSSKIEVILKNAIVNYSMYENFSTSWIRIPKYNFNSISLNN 960

QY 961 EYTIINCWNNNGWVSLNGEIIITWLTQTOEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
DB 961 EYTIINCWNNNGWVSLNGEIIITWLTQTOEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
QY 1081 EKEIKDLYDNQSGIILKDFWGDYLDYKPYMLNLYDPNKYVDVNNVGRGYMLKGP 1140
DB 1081 EKEIKDLYDNQSGIILKDFWGDYLDYKPYMLNLYDPNKYVDVNNVGRGYMLKGP 1140
QY 1141 GSVMTNLYNLSLYRGTKFIKKYASGNKDNIRVNRDVRVINVVKNGEYRLATNASQA 1200
DB 1141 GSVMTNLYNLSLYRGTKFIKKYASGNKDNIRVNRDVRVINVVKNGEYRLATNASQA 1200
QY 1201 GVEKILSALEIPDVGNIQSVMVMSKNDQGITNKCKWNLQDNGNDIGFTGFQFNNAK 1260
DB 1201 GVEKILSALEIPDVGNIQSVMVMSKNDQGITNKCKWNLQDNGNDIGFTGFQFNNAK 1260
QY 1261 LVASNNWNRQIERSRSLTGCSEWEPVDDGNGERPL 1296
DB 1261 LVASNNWNRQIERSRSLTGCSEWEPVDDGNGERPL 1296

RESULT 9

US-10-354-774-28
; Sequence 28, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/354,774
; FILING DATE: 30-Jan-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-354-774-28

Query Match 99.0%; Score 6756; DB 14; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVNKQFNYPKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIPERDTFNPESGDLN 60
DB 1 MQFVNKQFNYPKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIPIPERDTFNPESGDLN 60
QY 61 PPEAKQVPVSYVDSTYLTSDNEKNYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPEAKQVPVSYVDSTYLTSDNEKNYLKGVTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIOPDGSYRSEELNLVIIGPSADIIIOPECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIOPDGSYRSEELNLVIIGPSADIIIOPECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFEESLEVDNTPLLGAGKFPATDPAVTLAHELIIYAGHRLYGIAINPN 240
DB 181 GSTQYIRFSPDFTFGFEESLEVDNTPLLGAGKFPATDPAVTLAHELIIYAGHRLYGIAINPN 240
QY 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDLSQENEFLLYYNFKDIASTLNKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELRTFGGHDKAFIDLSQENEFLLYYNFKDIASTLNKA 300
QY 301 KSIIVGTTASLOQYMKNVFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
DB 301 KSIIVGTTASLOQYMKNVFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYTEDNFVKFKV 360
QY 361 LNRKTYLNFDKAVFKINIIVPKVNYTIYDGFNLRLNTLAAFNQONTTEINNMNFTKLKNFT 420
DB 361 LNRKTYLNFDKAVFKINIIVPKVNYTIYDGFNLRLNTLAAFNQONTTEINNMNFTKLKNFT 420
QY 421 GLFEFYKLLCVRGIIITSKTSKLDGYNKALNDLCIKVNWDLFPSPEDNFTNDLNGEE 480
DB 421 GLFEFYKLLCVRGIIITSKTSKLDGYNKALNDLCIKVNWDLFPSPEDNFTNDLNGEE 480
QY 481 ITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISIESSLSDIIIGOLELMPNIEFPNG 540
DB 481 ITSDTNIEAAEENISLDLIQOYYLTFNFDNEPENISIESSLSDIIIGOLELMPNIEFPNG 540
QY 541 KKYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVVTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTMFHYLRAQEPFHGKSRIALTNSVNEALLNPSRVVTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPALNIGNMLYKDDDFVGALIFSG 660
QY 661 AVILLEPIEPIAIPVLGTFFALVSIIANKVLTQVOTIDNALSCKRNEKWDDEVYKIVTNWLAK 720
DB 661 AVILLEPIEPIAIPVLGTFFALVSIIANKVLTQVOTIDNALSCKRNEKWDDEVYKIVTNWLAK 720
QY 721 VNTQIDILIRKMKKEALENOQAEATKAIINYOYQYTEEKKNINFNIDDLSSKLNESINKA 780
DB 721 VNTQIDILIRKMKKEALENOQAEATKAIINYOYQYTEEKKNINFNIDDLSSKLNESINKA 780
QY 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKIYDNRGTILIGVDRLKDK 840
DB 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKIYDNRGTILIGVDRLKDK 840
QY 841 VVNTLSTDIDPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRLYESNHLIDLSRYASIKINI 900
DB 841 VVNTLSTDIDPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRLYESNHLIDLSRYASIKINI 900
QY 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVVNSVNFSTFWIRIPKYFNSISLNN 960
DB 901 GSKVNFDPIDKNOIQLFNLESSKIEVLKNAIVVNSVNFSTFWIRIPKYFNSISLNN 960
QY 961 EYTIINCWNNNGWVSLNGEIIITWLTQTOEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
DB 961 EYTIINCWNNNGWVSLNGEIIITWLTQTOEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPIISNLGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080

Qy	1081	EKEIKOLYDNQNSGILKDPFGDYLDKDPKYMYMLNLYDPNKYVDVNNVYIRGIRGMYLKGPR	1140
Db	1081	EKEIKOLYDNQNSGILKDPFGDYLDKDPKYMYMLNLYDPNKYVDVNNVYIRGIRGMYLKGPR	1140
Qy	1141	GSVYTTNIIYLNSSIIYGTGTFIIKKYASGNKDNIVRNNDRYIVNVVKNKYEYRLATNASQA	1200
Db	1141	GSVYTTNIIYLNSSIIYGTGTFIIKKYASGNKDNIVRNNDRYIVNVVKNKYEYRLATNASQA	1200
Qy	1201	GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKKCNLQDNNGNDIGTGFHFQFNNAK	1260
Db	1201	GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKKCNLQDNNGNDIGTGFHFQFNNAK	1260
Qy	1261	LVASNNYNNROIERSRRTLGCWSWFIPIVDDGWSRPL	1296
Db	1261	LVASNNYNNROIERSRRTLGCWSWFIPIVDDGWSRPL	1296

RESULT 10

```

US-10-271-012-28
; Sequence 28, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; ; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; ; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-271-012-28

```

121	Qy	STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIIGPSADIIQOECKSGFGEVLNLTRNGY	180
121	Db	STIDTELKVIDTNCINVIQPDGSGYRSEELNLVIIIGPSADIIQOECKSGFGEVLNLTRNGY	180
181	Qy	GSTQYIRFSPDFTFGPEESLEVDVTNPLLGAGKFAFDPAVTLAHELIIYAGHRLYGIAINPN	240
181	Db	GSTQYIRFSPDFTFGPEESLEVDVTNPLLGAGKFAFDPAVTLAHELIIYAGHRLYGIAINPN	240
241	Qy	RVPKVNNTAIIYEMSGLEVSFEELRTPGGHDAKPIDSLQENFERLYYYNFKFOIASTLNKA	300
241	Db	RVPKVNNTAIIYEMSGLEVSFEELRTPGGHDAKPIDSLQENFERLYYYNFKFOIASTLNKA	300
301	Qy	KSTIVGTTASIQYMKNVFKPKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTDENFVKVFKV	360
301	Db	KSTIVGTTASIQYMKNVFKPKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTDENFVKVFKV	360
361	Qy	LNRKTYLNFDAKVFKINIVPKVNYTIIYDGFNLRTNTLAAANFNGQNTIEINNMFTKLKNFT	420
361	Db	LNRKTYLNFDAKVFKINIVPKVNYTIIYDGFNLRTNTLAAANFNGQNTIEINNMFTKLKNFT	420
421	Qy	GLPEFPYKLLICVRGIIITSKTKSLDKGVNKALNDLICIKVNRWDLPFSPSEDNFTNDLNKGBE	480
421	Db	GLPEFPYKLLICVRGIIITSKTKSLDKGVNKALNDLICIKVNRWDLPFSPSEDNFTNDLNKGBE	480
481	Qy	ITSDTNIEAAEENISLDLIIQQYYLTFNFONBENPENSIENTLSSDIIQOLELMPNIEIRFPNG	540
481	Db	ITSDTNIEAAEENISLDLIIQQYYLTFNFONBENPENSIENTLSSDIIQOLELMPNIEIRFPNG	540
541	Qy	KKYELDKYTMFYHLRAQEPHEGKSGRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA	600
541	Db	KKYELDKYTMFYHLRAQEPHEGKSGRIALTNSVNEALLNPSRVVTFPSSDYVKKVKNKATEA	600
601	Qy	AMPLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYGIPALNIGMLYKDDFVGALIPSG	660
601	Db	AMPLGWQEQLVYDFTDETSEVSTTDKIADITIIPIYGIPALNIGMLYKDDFVGALIPSG	660
661	Qy	AVILLEFIEPIAIPVLGTGFALVSYIANKVLTVQTDIDNALSKRNEKWDEYKVIIVTNWLAK	720
661	Db	AVILLEFIEPIAIPVLGTGFALVSYIANKVLTVQTDIDNALSKRNEKWDEYKVIIVTNWLAK	720
721	Qy	VNTQIDILRKQKKEALENOAEATKAIINTYQYQYTEEEKNNINFNIDDLSSKLNESINKA	780
721	Db	VNTQIDILRKQKKEALENOAEATKAIINTYQYQYTEEEKNNINFNIDDLSSKLNESINKA	780
781	Qy	MININKFLNQCSVSYLNMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDELKDK	840
781	Db	MININKFLNQCSVSYLNMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTLIGQVDELKDK	840
841	Qy	VNNTLSTDIPFQLSKYVDNQRLLSTPTEYIKNIINTSIIILNRYESNHLIDLSRYASKINI	900
841	Db	VNNTLSTDIPFQLSKYVDNQRLLSTPTEYIKNIINTSIIILNRYESNHLIDLSRYASKINI	900
901	Qy	GSKVNFDPDKNQIQULFNLESSEKIEVLKNAIYNYSMYENFSTSFWIRIPKYFNSISLNN	960
901	Db	GSKVNFDPDKNQIQULFNLESSEKIEVLKNAIYNYSMYENFSTSFWIRIPKYFNSISLNN	960
961	Qy	EYTIINCMMNNSGWKSVSLNYGBIIWTLQDTQBIKQRVVFKYSQMINISDIYINRWIFVTIT	1020
961	Db	EYTIINCMMNNSGWKSVSLNYGBIIWTLQDTQBIKQRVVFKYSQMINISDIYINRWIFVTIT	1020
1021	Qy	NNRLNNSKIYINGRLIDQKPISNLGNIIHASNNIMFKLDGCRDTHRYIWTIKYNLFPKELN	1080
1021	Db	NNRLNNSKIYINGRLIDQKPISNLGNIIHASNNIMFKLDGCRDTHRYIWTIKYNLFPKELN	1080
1081	Qy	EKEIKDLYDNQNSGILKDFWDGYLQYDKPYPYMLNLYDPNKYVDVNNVVGIRGYMLKGRP	1140
1081	Db	EKEIKDLYDNQNSGILKDFWDGYLQYDKPYPYMLNLYDPNKYVDVNNVVGIRGYMLKGRP	1140
1141	Qy	GSVWTTNIIYLNSSLRYGTFKPIIKKYASGNKONIVRNNDRVYINVVVYKNKEYRLATNASQA	1200
1141	Db	GSVWTTNIIYLNSSLRYGTFKPIIKKYASGNKONIVRNNDRVYINVVVYKNKEYRLATNASQA	1200

Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNNGNDIGFIGHQFNIAK 1260
Db 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNNGNDIGFIGHQFNIAK 1260
Qy 1261 LVASNNWNRQIERSRSLTGCSEWEPFVDDGGERPL 1296
Db 1261 LVASNNWNRQIERSRSLTGCSEWEPFVDDGGERPL 1296
RESULT 11
US-10-452-024-158
; Sequence 158, Application US/10452024
; Publication No. US20040013687A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Lance
; APPLICANT: Park, Jung-Beak
; APPLICANT: Maksymowich, Andrew
; TITLE OF INVENTION: Compositions and Methods For Transsepithelial Molecular Transport
; FILE REFERENCE: 9855-96U1
; CURRENT APPLICATION NUMBER: US/10/452,024
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: 60/384,949
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 158
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; US-10-452-024-158
Query Match 99.0%; Score 6756; DB 15; Length 1296;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MPFVNKQPNYKDPVNGVDIAYIKIPNAGQMPVAKPKHKNKIWIPIERDFTNPEEGDLN 60
Db 1 MQFVNKQPNYKDPVNGVDIAYIKIPNAGQMPVAKPKHKNKIWIPIERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYSDTYLSTDNEXDNLKGYTKLPERIYSDLGRMLTSLVRGIPFWGG 120
Db 61 PPEAKQVPVSYSDTYLSTDNEXDNLKGYTKLPERIYSDLGRMLTSLVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPGSYRSEBELNVIIGPSADIIQFECKSPGHEVLNTRNGY 180
Db 121 STIDTELKVIDTNCINVIQPGSYRSEBELNVIIGPSADIIQFECKSPGHEVLNTRNGY 180
Qy 181 GSTQYIRSPDPTFGFEESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 240
Db 181 GSTQYIRSPDPTFGFEESLEVDNPLLGAGKATDPATVLAHELIIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300
Db 241 RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFDIASTLNKA 300
Qy 301 KSIIVGTTASLOVMKNVFEKILLSDTSGKFSVDKLEDKLYKMLTEIYTDNDFVKFPKV 360
Db 301 KSIIVGTTASLOVMKNVFEKILLSDTSGKFSVDKLEDKLYKMLTEIYTDNDFVKFPKV 360
Qy 361 LNRKTYLNFDAKVPKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTENNNPFTLKNFT 420
Db 361 LNRKTYLNFDAKVPKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTENNNPFTLKNFT 420
Qy 421 GLFEFYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE 480
Db 421 GLFEFYKLLCVRGIIITSTKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE 480
Qy 481 ITSDTNIIEAABENISLDIIQYLYTTFNFDNBPENISLENLSSDIIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIIEAABENISLDIIQYLYTTFNFDNBPENISLENLSSDIIIGOLELMPNIEFPNG 540
Qy 541 KKYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600

Db 541 KKYELDKYTMFHYLRAQFEHGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDSTSVSTTKIADIITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
Db 601 AMFLGWVEQLVYDFTDSTSVSTTKIADIITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
Qy 661 AVILLEFIPBIAIPVLGTFALVSIANKVLTQTIDNALSKRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEFIPBIAIPVLGTFALVSIANKVLTQTIDNALSKRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENOQAEATKAIINYQNYQTEEEKNNININIDDLSSKNESINKA 780
Db 721 VNTQIDILIRKMKKEALENOQAEATKAIINYQNYQTEEEKNNININIDDLSSKNESINKA 780
Qy 781 MININKFLNOCVSYLNMNMIPIGVKRELEDFDASLKDALLKYIYDNRGTILIGQVDRDKK 840
Db 781 MININKFLNOCVSYLNMNMIPIGVKRELEDFDASLKDALLKYIYDNRGTILIGQVDRDKK 840
Qy 841 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900
Db 841 VNNLTSTDIPFQLSKYVDNQRLLSTFTTEYIKNIINTSILNRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKNOIQLEFNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKNOIQLEFNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCWNNNGSKVSLNYGEIITWLTQDTEIKORVVFYKYSOMINISDIYNRWIFVTIT 1020
Db 961 EYTIINCWNNNGSKVSLNYGEIITWLTQDTEIKORVVFYKYSOMINISDIYNRWIFVTIT 1020
Qy 1021 NNRLLNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
Db 1021 NNRLLNSKIYINGRLIDOKPISNLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
Qy 1081 EKEIKDIYDNQSNIGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNVGIRGYMYLKGPR 1140
Db 1081 EKEIKDIYDNQSNIGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNVGIRGYMYLKGPR 1140
Qy 1141 GSVMTNIIYLNSSLYRGTKFTIIKKYASGNKDNIVRNNDRVYINVVVKKEYRLATNASQA 1200
Db 1141 GSVMTNIIYLNSSLYRGTKFTIIKKYASGNKDNIVRNNDRVYINVVVKKEYRLATNASQA 1200
Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNNGNDIGFIGHQFNIAK 1260
Db 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKMNLQDNNNGNDIGFIGHQFNIAK 1260
Qy 1261 LVASNNWNRQIERSRSLTGCSEWEPFVDDGGERPL 1296
Db 1261 LVASNNWNRQIERSRSLTGCSEWEPFVDDGGERPL 1296

RESULT 12
US-10-729-122-28
; Sequence 28, Application US/10729122
; Publication No. US20040115215A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,122
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-729-122-28

Query Match          99.08; Score 6756; DB 16; Length 1296;
Best Local Similarity 99.84; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVKQPNYKDPVNGVDIAIKIPNAGQOPVKAQKIHNNKIWIPIPERDTFTNPEEGDLN 60
DB 1 MQFVKQPNYKDPVNGVDIAIKIPNAGQOPVKAQKIHNNKIWIPIPERDTFTNPEEGDLN 60
QY 61 PPPKQVPSYSDYSTYSLTDNEKDYLGKVTYKFERIYSTDIGRMLLTIVIRGIPFWGG 120
DB 61 PPPKQVPSYSDYSTYSLTDNEKDYLGKVTYKFERIYSTDIGRMLLTIVIRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIIOPECKSFGEHLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYRSEELNLIIGPSADIIIOPECKSFGEHLNLTNGY 180
QY 181 GSTQYIRFSDPTFGFPESLEVDNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
DB 181 GSTQYIRFSDPTFGFPESLEVDNPLLGAGKATDPAVTLAHELIVAGHRLYGIAPNP 240
QY 241 RVFKVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTANKA 300
DB 241 RVFKVNTNAYEMSGLEVSFEELTFGCHDAKFDISLQENEFRLYYNKKFKDIASLTANKA 300
QY 301 KSVIGTTASLQYMKVPEKYLISEDTSGKFSVDKLPDKLYKMLTEIYTEDNFVKPFKV 360
DB 301 KSVIGTTASLQYMKVPEKYLISEDTSGKFSVDKLPDKLYKMLTEIYTEDNFVKPFKV 360
QY 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNQONTNINNMTFKLNFT 420
DB 361 LNRKTYLNFDAKVPKINIVPKVNTIYDGFNLRNTNLAANFNQONTNINNMTFKLNFT 420
QY 421 GLPEFYKLLCVRGIIITSKTSLDKYGNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 480
DB 421 GLPEFYKLLCVRGIIITSKTSLDKYGNKALNDLCIKVNNWDLFPSPSEDNFTNDLNKGE 480
QY 481 ITSNTNTEAAEENISLDLIQOYLYTFNPDNEPENISLENSSDIIGOLELMPNIERFPNG 540
DB 481 ITSNTNTEAAEENISLDLIQOYLYTFNPDNEPENISLENSSDIIGOLELMPNIERFPNG 540
QY 541 KKYELDKYTFHFLYRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTFHFLYRAQEFHKGSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPNALNIGNMLYKDDFVGALIFSG 660
DB 601 AMFLGWVEQLVYDFTDSTSEVSTTDKIADITIIPIYGPNALNIGNMLYKDDFVGALIFSG 660
QY 661 AVILLEFPEIAIPVLGTALVSIAKVLTVOTIDNALSKRNEKWDVEVYKIYVNWLA 720
DB 661 AVILLEFPEIAIPVLGTALVSIAKVLTVOTIDNALSKRNEKWDVEVYKIYVNWLA 720

661 AVILLEFPEIAIPVLGTALVSIAKVLTVOTIDNALSKRNEKWDVEVYKIYVNWLA 720
721 VNTQIDILIRKMKKEALENAEATKAIINYQYNTQYTEEEKNNINFNIDDLSSKLNESINKA 780
721 VNTQIDILIRKMKKEALENAEATKAIINYQYNTQYTEEEKNNINFNIDDLSSKLNESINKA 780
781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALLKIYDNRGTLIGQVDRUKDK 840
781 MININKFLNQCYSVYLMNSMIPYGVKRLDFDASLKDALLKIYDNRGTLIGQVDRUKDK 840
841 VNTLSTDIIPFQLSKVVDNRLLSTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
841 VNTLSTDIIPFQLSKVVDNRLLSTTEYIKNIINTSILNLRYESNHLIDLSRYASKINI 900
901 GSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVYNSMYENFSTFWIRIPKYNFNSISLNN 960
901 GSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVYNSMYENFSTFWIRIPKYNFNSISLNN 960
961 EYTIINCMMNSGKVSILNAYGEIITWLODQEIQRVVFYKYSOMINISDIYINRWIFVTIT 1020
961 EYTIINCMMNSGKVSILNAYGEIITWLODQEIQRVVFYKYSOMINISDIYINRWIFVTIT 1020
1021 NRELNNKIYINGRLIDQKIPISNLGNIHASNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
1021 NRELNNKIYINGRLIDQKIPISNLGNIHASNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
1081 EKEIKDYLNQNSGILKDFWGDYLDQYDFPYMLNLYDPNKYVDVNNVGIRGYMLKGP 1140
1081 EKEIKDYLNQNSGILKDFWGDYLDQYDFPYMLNLYDPNKYVDVNNVGIRGYMLKGP 1140
1141 GSVMTTNIYLNSSLYRGTFIINKYASGKNKDNIVRNDRVYINNVVKNKEYRLATNASQA 1200
1141 GSVMTTNIYLNSSLYRGTFIINKYASGKNKDNIVRNDRVYINNVVKNKEYRLATNASQA 1200
1201 GVEKILSALIEIPDVGNLISQVVMKSKNDQGITNKKQNLQDNNNDIGFIFGHQFNNAK 1260
1201 GVEKILSALIEIPDVGNLISQVVMKSKNDQGITNKKQNLQDNNNDIGFIFGHQFNNAK 1260
1261 LVASNNYNRQIERSRRTLCGSWEFIPVDGWSGERPL 1296
1261 LVASNNYNRQIERSRRTLCGSWEFIPVDGWSGERPL 1296

RESULT 13
US-10-729-039-28
; Sequence 28, Application US/10729039
; Publication No. US20040142455A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Attorney: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/729,039
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: 1996/08/28
; ATTORNEY/AGENT INFORMATION:
```

```

;
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40, 027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-729-039-28

Query Match      99.08; Score 6756; DB 16; Length 1296;
Best Local Similarity 99.84; Pred. No. 0;
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPFVVKQPNYKDPVNGVDIAIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPBEGDLN 60
DB 1 MQFVVKQPNYKDPVNGVDIAIKIPNAGQMPVKAFKIHKNKIWIPIPERDFTTNPBEGDLN 60
QY 61 PPPEAKQPVSYDYSTYLSDTNEKDNKYLKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
DB 61 PPPEAKQPVSYDYSTYLSDTNEKDNKYLKGYTKLPERIYSTDLGRMLLTSIVRGIPFWGG 120
QY 121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQPECKSFGEHVLNLTNGY 180
DB 121 STIDTELKVIDTNCINVIQDGSYRSBELNLVIIGPSADIIQPECKSFGEHVLNLTNGY 180
QY 181 GSTQYIRFSPDFTFGFESLEVDNPLLAGKAFATDPAVTIAHELIIYAGHRLYGIAINPN 240
DB 181 GSTQYIRFSPDFTFGFESLEVDNPLLAGKAFATDPAVTIAHELIIYAGHRLYGIAINPN 240
QY 241 RVFKVNTNAYEMSGLEVSPELFTFGHDHAKFTDSLOENEFRLYYNKKFDIATSLINKA 300
DB 241 RVFKVNTNAYEMSGLEVSPELFTFGHDHAKFTDSLOENEFRLYYNKKFDIATSLINKA 300
QY 301 KSIYVGTASLOYKMNVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTDNDFVKKFKV 360
DB 301 KSIYVGTASLOYKMNVPKFKYLLSEDTSGKFSVDKLFKDKLYKMLTEIYTDNDFVKKFKV 360
QY 361 LNRKTYLNFDFKAVFKINIPKVNTIYDGFNLRLNLAANFNGQNTENNMFVKLKNFT 420
DB 361 LNRKTYLNFDFKAVFKINIPKVNTIYDGFNLRLNLAANFNGQNTENNMFVKLKNFT 420
QY 421 GLFEPYKLLCVRGITTSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
DB 421 GLFEPYKLLCVRGITTSKTSKLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTDLNKGEE 480
QY 481 ITSDTNEAARENISLDLIQYLYLTNFDPNENISLENSSDIIGOLELMPNTERFPNG 540
DB 481 ITSDTNEAARENISLDLIQYLYLTNFDPNENISLENSSDIIGOLELMPNTERFPNG 540
QY 541 KKYELDKYTMFHYLRAQFEHGKSRIALTNSVEALLNPSRVYTFSSDYVKKVKNKATEA 600
DB 541 KKYELDKYTMFHYLRAQFEHGKSRIALTNSVEALLNPSRVYTFSSDYVKKVKNKATEA 600
QY 601 AMFLGWVQLVYDFDTESEVSTTDKIADITIIPIYIGPALNIGNMLYKDPFVGLIFSG 660
DB 601 AMFLGWVQLVYDFDTESEVSTTDKIADITIIPIYIGPALNIGNMLYKDPFVGLIFSG 660
QY 661 AVILLEFPIETAIPLVGLFALVSYTANKVLTQVTDIDNALSKEKWEDEVYKVIYTNMLAK 720
DB 661 AVILLEFPIETAIPLVGLFALVSYTANKVLTQVTDIDNALSKEKWEDEVYKVIYTNMLAK 720
QY 721 VNTQIDTLIRKKMKEALENOAETAKAIINYQNYTEEEKNNINFDLSSKLNESINKA 780
DB 721 VNTQIDTLIRKKMKEALENOAETAKAIINYQNYTEEEKNNINFDLSSKLNESINKA 780
QY 781 MININKFLNOCSSVYLMNSMIPYGVKRLDFDASLKDALLYVDNRGTLIGQVDRDKDK 840
DB 781 MININKFLNOCSSVYLMNSMIPYGVKRLDFDASLKDALLYVDNRGTLIGQVDRDKDK 840
QY 841 VVNTLSTDIPIFQLSKYVDNQRLLSITFTEYIKNIINTSIILNRYESNHLIDLRYASKINI 900
DB 841 VVNTLSTDIPIFQLSKYVDNQRLLSITFTEYIKNIINTSIILNRYESNHLIDLRYASKINI 900
QY 901 GSKVNPDPIDKQIQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWRIPKYPFNSISLNN 960
DB 901 GSKVNPDPIDKQIQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWRIPKYPFNSISLNN 960
QY 961 EYTIINCMMNSGKVSLSNYGELIITWLTQTEIKQVWVFKYSQMINISDIYNRWIFVTIT 1020
DB 961 EYTIINCMMNSGKVSLSNYGELIITWLTQTEIKQVWVFKYSQMINISDIYNRWIFVTIT 1020
QY 1021 NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMEKLDGCRDTHRYIWIYKYNLFDKELN 1080
DB 1021 NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMEKLDGCRDTHRYIWIYKYNLFDKELN 1080
QY 1081 EKEIKDLYDNQNSGILKDFWGDYLOQDKPYMYMLNLYDNKYVDVNNVGIKRGVYMLKGR 1140
DB 1081 EKEIKDLYDNQNSGILKDFWGDYLOQDKPYMYMLNLYDNKYVDVNNVGIKRGVYMLKGR 1140
QY 1141 GSVMTTNIYLNSSLYRGTFKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
DB 1141 GSVMTTNIYLNSSLYRGTFKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
QY 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKCQKQNDQNDGNDIGIFGHQFNNAK 1260
DB 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKCQKQNDQNDGNDIGIFGHQFNNAK 1260
QY 1261 LVASNNYRQIERSSRTLCGSWEFIPIVDDGWSRPL 1296
DB 1261 LVASNNYRQIERSSRTLCGSWEFIPIVDDGWSRPL 1296

RESULT 14
US-10-729-527-28
; Sequence 28, Application US/10729527
; Publication No. US20040219637A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thallev, Bruce S.
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10729,527
; FILING DATE: 05-Dec-2003
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40, 027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
```

;	;	LENGTH: 1296 amino acids			
;	;	TYPE: amino acid			
;	;	TOPOLOGY: linear			
;	;	MOLECULE TYPE: protein			
;	;	SEQUENCE DESCRIPTION: SEQ ID NO: 28:			
US-10-729-527-28					
Query Match 99.0%; Score 6756; DB 17; Length 1296;					
Best Local Similarity 99.8%; Pred. No. 0;					
Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;					
QY	1	MPFVKQFNKDPVNGVDIAVIKIPNAGQMPVKAFLHNKIWIPIPERDFTNPEEGDLN	60		
DB	1	MQFVKQFNKDPVNGVDIAVIKIPNAGQMPVKAFLHNKIWIPIPERDFTNPEEGDLN	60		
QY	61	PPPEAKQPVSYDYDSTYLSQDNEKDYKLVKLFERIYSTDLGRMLLTISVIRGIPWGG	120		
DB	61	PPPEAKQPVSYDYDSTYLSQDNEKDYKLVKLVKLFERIYSTDLGRMLLTISVIRGIPWGG	120		
QY	121	STIDTELKVIDTNCINVIQDGSVRSEBELNVIIGPSADIIQPECKSFGEHVLNLTNGY	180		
DB	121	STIDTELKVIDTNCINVIQDGSVRSEBELNVIIGPSADIIQPECKSFGEHVLNLTNGY	180		
QY	181	GSTQYIRSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP	240		
DB	181	GSTQYIRSPDFTFGFESLEVDNPLLGAGKATDPAVTLAHELIIYAGHRLYGIAPNP	240		
QY	241	RVFKVNTNAYEMSGLEVSPEELTFGCHDAKFTDSLQENEFRLYYNKPDKIASTLNKA	300		
DB	241	RVFKVNTNAYEMSGLEVSPEELTFGCHDAKFTDSLQENEFRLYYNKPDKIASTLNKA	300		
QY	301	KSIIVGTTASLOYQMNVPKFKYLLSEDTSGKFSVDKLPKLYKMLTEIYEDNFPVKPKV	360		
DB	301	KSIIVGTTASLOYQMNVPKFKYLLSEDTSGKFSVDKLPKLYKMLTEIYEDNFPVKPKV	360		
QY	361	LNKRTYLNFKKAVKINIVPKVNTIYDGNLRTNLAANFGQNTINNWNFTKLNFT	420		
DB	361	LNKRTYLNFKKAVKINIVPKVNTIYDGNLRTNLAANFGQNTINNWNFTKLNFT	420		
QY	421	GLPEFYKLLCVRGIIITSKLSLDGYNKALNDLCIKVNNWDLFPSPEDNFTDNLKGE	480		
DB	421	GLPEFYKLLCVRGIIITSKLSLDGYNKALNDLCIKVNNWDLFPSPEDNFTDNLKGE	480		
QY	481	ITSDNTIEAABENISLDLIIQYYLTFNFDNENPENSIENTSSDIIGOLELMPNIERPNG	540		
DB	481	ITSDNTIEAABENISLDLIIQYYLTFNFDNENPENSIENTSSDIIGOLELMPNIERPNG	540		
QY	541	KYELDKYTWPHYLRAQEFPHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600		
DB	541	KYELDKYTWPHYLRAQEFPHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKVKNKATEA	600		
QY	601	AMFLGWEOQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	660		
DB	601	AMFLGWEOQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG	660		
QY	661	AVILLEFIPEIAPVLGTGTFALVSIAKVLTVQTDIDNALSKEKDEVKYKIVTNWLAK	720		
DB	661	AVILLEFIPEIAPVLGTGTFALVSIAKVLTVQTDIDNALSKEKDEVKYKIVTNWLAK	720		
QY	721	VNTQIDILIRKQKALENQAETKAIINYNQVQTEBEKNININIDDLSSKLNESINKA	780		
DB	721	VNTQIDILIRKQKALENQAETKAIINYNQVQTEBEKNININIDDLSSKLNESINKA	780		
QY	781	MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALIKYIYDNRGTGIGQVDRUKDK	840		
DB	781	MININKFLNOCVSYLMNSMIPYGVKLEDFDASLKDALIKYIYDNRGTGIGQVDRUKDK	840		
QY	841	VNNTLSTDIIPQLSKYVDNORLLSTFTYEYKNIINTSILNRYESNHLIDLRYASKINI	900		
DB	841	VNNTLSTDIIPQLSKYVDNORLLSTFTYEYKNIINTSILNRYESNHLIDLRYASKINI	900		
QY	901	GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVTNSMYENFSTFWIRIPKYNFNSISLNN	960		

Db	901	GSKVNFDPIDKQIQLFNLESSKIEVLKNAIVTNSMYENFSTFWIRIPKYNFNSISLNN	960		
QY	961	EYTIINCMMNSGWKVSILNYGIIWTLQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020		
Db	961	EYTIINCMMNSGWKVSILNYGIIWTLQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020		
QY	1021	NNRLNNSKIYINGRLIDOKPISNLGNIHASNNIMPKLDGCRDTHRYIWKIFNLPDKELN	1080		
Db	1021	NNRLNNSKIYINGRLIDOKPISNLGNIHASNNIMPKLDGCRDTHRYIWKIFNLPDKELN	1080		
QY	1081	EKEIKDLYDNQNSGILKDFWGDYLDQYDKPYPMYMLNLYDPNKYVDVNNVGIRGYMYLKQPR	1140		
Db	1081	EKEIKDLYDNQNSGILKDFWGDYLDQYDKPYPMYMLNLYDPNKYVDVNNVGIRGYMYLKQPR	1140		
QY	1141	GSVMTTNIYLNSSLYRGTKFIKKYASGNKONIVRNDRVYINVVVKNKYRLATNASQA	1200		
Db	1141	GSVMTTNIYLNSSLYRGTKFIKKYASGNKONIVRNDRVYINVVVKNKYRLATNASQA	1200		
QY	1201	GVEKILSALEIPDVGNLSQVVMVMSKNDQGTNCKMNLQDNGNDIGFIGHQFNNIAK	1260		
Db	1201	GVEKILSALEIPDVGNLSQVVMVMSKNDQGTNCKMNLQDNGNDIGFIGHQFNNIAK	1260		
QY	1261	LVASWYNRQIRERSRTILGCSWEFIPVDGHWGERPL	1296		
Db	1261	LVASWYNRQIRERSRTILGCSWEFIPVDGHWGERPL	1296		
RESULT 15					
US-10-727-898-28					
; Sequence 28, Application US/10727898					
; Publication No. US20040235118A1					
; GENERAL INFORMATION:					
; APPLICANT: Walliams, James A.					
; Thallev, Bruce S.					
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium					
; Botulinum Neurotoxin					
; NUMBER OF SEQUENCES: 82					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: Medlen & Carroll					
; STREET: 220 Montgomery Street, Suite 2200					
; CITY: San Francisco					
; STATE: California					
; COUNTRY: United States of America					
; ZIP: 94104					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: Patent In Release #1.0, Version #1.30					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/10/727,898					
; FILING DATE: 04-Dec-2003					
; CLASSIFICATION: 424					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US/08/704,159					
; FILING DATE: <Unknown>					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Ingolia, Diane E.					
; REGISTRATION NUMBER: 40,027					
; REFERENCE/DOCKET NUMBER: OPHD-02304					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: (415) 705-8410					
; TELEFAX: (415) 397-8338					
; INFORMATION FOR SEQ ID NO: 28:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 1296 amino acids					
; TYPE: amino acid					
; TOPOLOGY: linear					
; MOLECULE TYPE: protein					
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:					
US-10-727-898-28					
Query Match 99.0%; Score 6756; DB 17; Length 1296;					

Query Match

99.0%; Score 6756; DB 17; Length 1296;

Best Local Similarity 99.8%; Pred. No. 0; Matches 1293; Conservative 1; Mismatches 2; Indels 0; Gaps 0;									
Qy	1	MPFVNKQFNKDPVNGVDIAIYKIPNAGOMQPVKAFKHNKIWIIPERDFTFTPEEGDLN	60						
Db	1	MQFVNKQFNKDPVNGVDIAIYKIPNAGOMQPVKAFKHNKIWIIPERDFTFTPEEGDLN	60						
Qy	61	PPPEAKQVPVSYDSTYSLTDNEKDNVKGVTKLPERIYSTDLGRMLLTISVIGIPFWGG	120						
Db	61	PPPEAKQVPVSYDSTYSLTDNEKDNVKGVTKLPERIYSTDLGRMLLTISVIGIPFWGG	120						
Qy	121	STIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIQPECKSFGEHVLNLTNGY	180						
Db	121	STIDTELKVIDTNCINVIQPDGSYRSEELNLIIGPSADIQPECKSFGEHVLNLTNGY	180						
Qy	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAKFATPAVTLAHLIYAGHRLVGIAPNP	240						
Db	181	GSTQYIRFSPDFTFGFESLEVDNPLLGAKFATPAVTLAHLIYAGHRLVGIAPNP	240						
Qy	241	RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFTDSIQENEFRLYYYNKFCDIASTLKA	300						
Db	241	RVFKVNTNAYYEMSGLEVSPEELRTFGHDAKFTDSIQENEFRLYYYNKFCDIASTLKA	300						
Qy	301	KSIIVGTTASLQYMKVFKPKYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKPKV	360						
Db	301	KSIIVGTTASLQYMKVFKPKYLLSEDTSGKFSVDKLFKLYKMLTEIYTEDNPFVKPKV	360						
Qy	361	LNRKTYLNFDAVKFKNIVPKVNTIYDGNLRLNTLAAFPNGONTNINNNFTKLKNT	420						
Db	361	LNRKTYLNFDAVKFKNIVPKVNTIYDGNLRLNTLAAFPNGONTNINNNFTKLKNT	420						
Qy	421	GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDFTNDLNKGE	480						
Db	421	GLFEFYKLLCVRGIIITSKTSLDKGYNKALNDLCIKVNNWDLFPSPSEDFTNDLNKGE	480						
Qy	481	ITSDTNIEAABENTSLDIIQOYIYITFNPDPENISIENTSSDIIGOLELMPNIERPNG	540						
Db	481	ITSDTNIEAABENTSLDIIQOYIYITFNPDPENISIENTSSDIIGOLELMPNIERPNG	540						
Qy	541	KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDVVKVKNKATEA	600						
Db	541	KKYELDKYTMFHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDVVKVKNKATEA	600						
Qy	601	AMFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKODFVGALIFSG	660						
Db	601	AMFLGWVQLVYDFDETSEVSTTDKIADITIIPIYIPALNIGNMLYKODFVGALIFSG	660						
Qy	661	AVILLEFIPEIAIPVLGTFALVSYIANKVLTVOIDNALSKRNEKWDVYKIIYVTNWLAK	720						
Db	661	AVILLEFIPEIAIPVLGTFALVSYIANKVLTVOIDNALSKRNEKWDVYKIIYVTNWLAK	720						
Qy	721	VNTQIDLLRKKWKEALENQAEATKAIINYQVQYTEBEKNINFNIDDLSSKLNESINKA	780						
Db	721	VNTQIDLLRKKWKEALENQAEATKAIINYQVQYTEBEKNINFNIDDLSSKLNESINKA	780						
Qy	781	MININKFLNQCYSVSLNMSIPIYGKRLDFDASLKDALLKXIYDNRGTLLIGQVDRLKDK	840						
Db	781	MININKFLNQCYSVSLNMSIPIYGKRLDFDASLKDALLKXIYDNRGTLLIGQVDRLKDK	840						
Qy	841	VNNTLSTDIIPOLSKYVDNORLLSTFTYIYKNIINTSILNLRYESNHLIDLRYASKINI	900						
Db	841	VNNTLSTDIIPOLSKYVDNORLLSTFTYIYKNIINTSILNLRYESNHLIDLRYASKINI	900						
Qy	901	GSKVNFDPIDKNOIQOLFNESSKBEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960						
Db	901	GSKVNFDPIDKNOIQOLFNESSKBEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN	960						
Qy	961	EYTIINCINNENSGWKVSLNYGEIITWLTQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020						
Db	961	EYTIINCINNENSGWKVSLNYGEIITWLTQDTQEIQRVVFKYSQMINISDYINRWIFVTIT	1020						
Qy	1021	NNRLNNSKIYINGRLIDQKPISNLGNIIHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN	1080						

Search completed: January 31, 2005, 14:12:58
Job time : 181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 52 Seconds
(without alignments)
2409.120 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNKYDVPNGVDIA.....EFIPVDGNGERPLHHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6768	99.2	1296	1 BFLCLAB	bontoxilysin (EC 3
2	6158	90.3	1296	2 I40645	botulinum neurotox
3	2504.5	36.7	1268	2 S23411	botulinum neurotox
4	2339	34.3	1252	2 S21178	botulinum neurotox
5	2324.5	34.1	1251	2 JH0256	botulinum neurotox
6	2296.5	33.7	1274	2 I40813	botulinum neurotox
7	2285	33.5	1297	2 S39791	neurotoxin - Clost
8	2271	33.3	1291	1 A48940	bontoxilysin (EC 3
9	2267	33.2	1291	2 I40631	non-proteolytic bo
10	1892.5	27.7	1315	1 BFLCLAB	bontoxilysin (EC 3
11	1890.5	27.7	1276	2 S11455	botulinum neurotox
12	1851	27.1	1285	2 S70582	botulinum neurotox
13	1771	26.0	1291	2 A49777	botulinum neurotox
14	1771	26.0	1291	2 S46431	botulinum neurotox
15	847	12.4	366	2 S48110	neurotoxin type F
16	805.5	11.8	367	2 S48106	neurotoxin type E
17	804.5	11.8	369	2 S48109	neurotoxin type F
18	616	9.0	552	2 A49928	neurotoxin Cl, BoN
19	614.5	9.0	1162	2 A47708	progenitor toxin n
20	614.5	9.0	1162	2 I40817	botulinum toxin no
21	583	8.5	1193	2 S68218	botulinum neurotox
22	567.5	8.3	1196	2 J01467	toxin, nontoxic co
23	566.5	8.3	1196	2 S64430	botulinum neurotox
24	553.5	8.1	1193	2 J04901	nontoxic-nonhemagg
25	534.5	7.8	1165	2 I40644	botulinum neurotox
26	318.5	4.7	2401	2 T28676	roptory protein -
27	310	4.5	960	2 S72284	DNA-directed RNA p
28	302.5	4.4	2269	2 T28677	roptory protein -
29	297.5	4.4	1127	2 T28317	ORF MSV156 hypothe

30	280	4.1	3724	2	T18427	hypothetical prote
31	280	4.1	4550	2	T18440	hypothetical prote
32	276.5	4.1	2485	1	H71621	serine/threonine-s
33	276.5	4.1	4688	2	F82885	hypothetical prote
34	272.5	4.0	1830	2	E82909	conserved hypotet
35	268	3.9	3394	2	T18501	hypothetical prote
36	267	3.9	2819	2	A90551	conserved hypotet
37	265	3.9	2380	2	E71604	hypothetical prote
38	264.5	3.9	4981	2	T18489	hypothetical prote
39	264	3.9	1183	2	F90559	conserved hypotet
40	261.5	3.8	2539	2	B71619	hypothetical prote
41	259.5	3.8	1121	2	F71613	hypothetical prote
42	258.5	3.8	1447	2	F82909	hypothetical prote
43	257.5	3.8	1346	2	G71613	hypothetical prote
44	257	3.8	5005	2	F82884	hypothetical prote
45	256.5	3.8	2136	2	A05037	hypothetical prote

ALIGNMENTS

RESULT 1

BTCLAB

bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum

N;Alternate names: botulinum neurotoxin type A

C;Species: Clostridium botulinum

C;Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 09-Jul-2004

C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000

R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.

J. Biol. Chem. 265, 9153-9158, 1990

A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other

A;Reference number: A35294; MUID:90264400; PMID:2160960

A;Accession: A35294

A;Molecule type: DNA

A;Residues: 1-1296 <BIN>

A;Cross-references: UNIPROT:P10845; GB:M30196; NID:G144864; PIDN:AAA23262.1; PID:G144865

A;Experimental source: strain 62A, subtype A

R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.

Eur. J. Biochem. 189, 73-81, 1990

A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin

A;Reference number: S09492; MUID:90235864; PMID:2185020

A;Accession: S09492

A;Molecule type: DNA

A;Residues: 1, 'Q', 3-26, 'V', 28-1296 <THO>

A;Cross-references: EMBL:X52066; NID:G40381; PIDN:CAA36289.1; PID:G40382

A;Experimental source: NCTC 2916

R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.

FEBS Lett. 376, 41-44, 1995

A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components

A;Reference number: S67988; MUID:96096783; PMID:8521962

A;Accession: S68220

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-12 <FLU>

A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:G2160224

R;Betley, M.J.; Somers, E.; DasGupta, B.R.

Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989

A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term

A;Reference number: A33401; MUID:89350959; PMID:2669749

A;Accession: A33401

A;Molecule type: DNA

A;Residues: 1-35 <BET>

A;Cross-references: GB:M27892; NID:G144880; PIDN:AAA23269.1; PID:G551776

R;Gimenez, J.A.; DasGupta, B.R.

J. Protein Chem. 12, 351-363, 1993

A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an

A;Reference number: A53884; MUID:94000342; PMID:8397993

A;Accession: A53884

A;Status: preliminary

A;Molecule type: protein

A;Residues: 867-880:1148-1217 'Y', 1219 <GIM>

A;Experimental source: strain Hall

A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex

R:DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A:Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and around the active site
A:Reference number: A60025; MUID:91120847; PMID:2126206
A:Accession: A60025
A:Molecule type: protein
A:Residues: 2-6;445-453, 'X', 455-457 <DAS1>
R:DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A:Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A:Reference number: A27000
A:Accession: A27000
A:Molecule type: protein
A:Residues: 2-47 <DAS2>
R:Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; J. Biol. Chem. 269, 1617-1620, 1994
A:Title: Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.
A:Reference number: A49708; MUID:94124495; PMID:8294407
A:Contents: annotation
C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses
C:Genetics:
A:Gene: atx; bota
C:Function:
A:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C:Superfamily: tetanus toxin
C:Keywords: disulfide bond; hydrolase; metalloprotease; neurotoxin; transmembrane protein
F:2-444/Product: botulinum A light chain #status experimental <LIGHT>
F:445-1296/Product: botulinum A heavy chain #status experimental <HY>
F:223,227/Binding site: zinc (His) #status predicted
F:224/Active site: Glu #status predicted

Query Match 99.2%; Score 6768; DB 1; Length 1296;
Best Local Similarity 99.9%; Pred. No. 1.7e-287;
Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MPFVNQPNQKDPVNGVDIAIKIPNAGQMPVKAFLHKKIWIIPERDFTNPEEGDLN 60
Db 1 MPFVNQPNQKDPVNGVDIAIKIPNAGQMPVKAFLHKKIWIIPERDFTNPEEGDLN 60
Qy 61 PPEAKQVPVSYDSTYLTSTNEKNDYLGKVTLPERYSTDLGRMLLTSLVRGIPFWGG 120
Db 61 PPEAKQVPVSYDSTYLTSTNEKNDYLGKVTLPERYSTDLGRMLLTSLVRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSFGEVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQDGSYRSEELNVIIGPSADIIQFECKSFGEVLNLTNGY 180
Qy 181 GSTQYIRSPDFTFGFESLSEVONPLLGAKGFATDPVTLAHLIYAGHRLYGIAPNP 240
Db 181 GSTQYIRSPDFTFGFESLSEVONPLLGAKGFATDPVTLAHLIYAGHRLYGIAPNP 240
Qy 241 RVFKNTNAYYEMSGLEVSFEELRTFGGDAKFDISLQENEFRLYYNKPDKDIASLTNKA 300
Db 241 RVFKNTNAYYEMSGLEVSFEELRTFGGDAKFDISLQENEFRLYYNKPDKDIASLTNKA 300
Qy 301 KSIIVGTTASLQVMKNVFKKYLSDTSCKFSVDKLFKDKLYKMLTEIYTDNEDNVKPKFV 360
Db 301 KSIIVGTTASLQVMKNVFKKYLSDTSCKFSVDKLFKDKLYKMLTEIYTDNEDNVKPKFV 360
Qy 361 LNRKTYLNFDKAVFKINIVPKVNTIYDGFNLNRTNLAANFGQNTENNNANFKLKNFT 420
Db 361 LNRKTYLNFDKAVFKINIVPKVNTIYDGFNLNRTNLAANFGQNTENNNANFKLKNFT 420
Qy 421 GLFEYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE 480
Db 421 GLFEYKLLCVRGIIITSKSLDKGYNKALNDLCIKVNNWDLFFSPSDNFTNDLNKGE 480
Qy 481 ITSDTNIAAENISLDLIQOYITFTNFDNBPENISLENLSDIIQOLELMPNIEFPNG 540
Db 481 ITSDTNIAAENISLDLIQOYITFTNFDNBPENISLENLSDIIQOLELMPNIEFPNG 540
Qy 541 KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFSSDYVKKNKATEA 600
Db 541 KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFSSDYVKKNKATEA 600

Db 541 KYELDKYTMFHYLRAQFEHGKSRIALTNSVNEALLNPSRVVTFSSDYVKKNKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYIGPALNIGNMLYKDDFVGALIFSG 660
Qy 661 AVILLEFIEPIAIPVLGTALVSIYANKVLTVQTIIDNALSQRNEKWDVYKIVTNWLAK 720
Db 661 AVILLEFIEPIAIPVLGTALVSIYANKVLTVQTIIDNALSQRNEKWDVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOYTEBEKKNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAEATKAIINYOYNOYTEBEKKNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCSSYVLMNSMIPYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRLKDK 840
Db 781 MININKFLNOCSSYVLMNSMIPYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRLKDK 840
Qy 841 VNNLTSTDIPFQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Db 841 VNNLTSTDIPFQLSKYVDNQRLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKNOIQIOLFNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKNOIQIOLFNLESSKIEVILKNAIVNMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCMMNSGKWSYVLMNSMIPYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRLKDK 1020
Db 961 EYTIINCMMNSGKWSYVLMNSMIPYGVKRLDFDASLKDALLKIYIDNRGTILIGQVDRLKDK 1020
Qy 1021 NNRLLNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYKFLDKELN 1080
Db 1021 NNRLLNSKIYINGRLIDQKPISNLGNHASNIMFKLDGCRDTHRYIWIYKFLDKELN 1080
Qy 1081 EKEIKDLYDNQNSGILKDPFGDYLDQKPYVLMNLYDPNKKYDVNNGVIRGYWYLGKPR 1140
Db 1081 EKEIKDLYDNQNSGILKDPFGDYLDQKPYVLMNLYDPNKKYDVNNGVIRGYWYLGKPR 1140
Qy 1141 GSVMTNTIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Db 1141 GSVMTNTIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRYIVNVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKANLQDNNGNDIGFGFHQFNIAK 1260
Db 1201 GVEKILSALBIPDVGNLSQVVMKSKNDQGITNCKANLQDNNGNDIGFGFHQFNIAK 1260
Qy 1261 LVASNNVNRQIERSSRTLGCSEWFIPIVDDGWERPL 1296
Db 1261 LVASNNVNRQIERSSRTLGCSEWFIPIVDDGWERPL 1296

RESULT 2

140645
botulinum neurotoxin type A - Clostridium botulinum

C:Species: Clostridium botulinum

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C:Accession: I40645

R:Willlems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.

Res. Microbiol. 144, 547-556, 1993

A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A

A:Reference number: 140645; MUID:94143603; PMID:8310180

A:Accession: I40645

A:Status: preliminary; translated from GE/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1296 <RES>

A:Cross-references: UNIPROT.Q45894; EMBL.X73423; NID:9507070; PIDN:CAA51824.1; PID:9507070

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 90.3%; Score 6158; DB 2; Length 1296;

Best Local Similarity 89.9%; Pred. No. 6.4e-261;

Matches 1165; Conservative 66; Mismatches 65; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEBGDLN 60
Db 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEBGDLN 60
Qy 61 PPEAKQVPVSYDYDSTYLSKDNEKNYLKGVTKLPERIYSTDLGRMLLTISVIRGIPFWGG 120
Db 61 PPEAKQVPVSYDYDSTYLSKDNEKNYLKGVTKLPERIYSTDLGRMLLTISVIRGIPFWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSYRSEBELNLIIGPSADIIQFECKSFHGVNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSYRSEBELNLIIGPSADIIQFECKSFHGVNLTNGY 180
Qy 181 GSTQYIRFSPDFTGFPESLEVDNPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTGFPESLEVDNPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Qy 241 RVFKVNTNAYEMSGLEVSFEELRTFGHDAKPIIDSIQENEFRLYYNKPDKDASTLANKA 300
Db 241 RVFKVNTNAYEMSGLEVSFEELRTFGHDAKPIIDSIQENEFRLYYNKPDKDASTLANKA 300
Qy 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Db 301 KSIIVGTTASLQYMKVNFKEKYLLEDTSKGFSVDKLFKDKLYKMLTEIYTEDNFVKPFKV 360
Qy 361 LNRKTYLNFQKAVPKINI VPKVNTIYDGNLRLNLTNLAANFGONTENNMTFKLNFT 420
Db 361 LNRKTYLNFQKAVPKINI VPKVNTIYDGNLRLNLTNLAANFGONTENNMTFKLNFT 420
Qy 421 GLPEFYKLCVRGIIITKTSKSLDGYNKALNDLCIKUNWDLFPSPEDNFTDLNKGEE 480
Db 421 GLPEFYKLCVRGIIITKTSKSLDGYNKALNDLCIKUNWDLFPSPEDNFTDLNKGEE 480
Qy 481 ITSNTIEAEEENISLDLIQOYLYTTFDNEPENISIENLSSDIIQGLELMPNIERPNG 540
Db 481 ITSNTIEAEEENISLDLIQOYLYTTFDNEPENISIENLSSDIIQGLELMPNIERPNG 540
Qy 541 KKYELDKYTHFYLRQAEPFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKKVKNKATEA 600
Db 541 KKYELDKYTHFYLRQAEPFHGKSRIALTNSVNEALLNPSRVYTFPSSDYVKKVKNKATEA 600
Qy 601 AMELGWEOQVYDFTDETSVSTDKIADITIIIPYIGPALNIGMILYKDDFCALIFSG 660
Db 601 AMELGWEOQVYDFTDETSVSTDKIADITIIIPYIGPALNIGMILYKDDFCALIFSG 660
Qy 661 AVILLEIPEIAPVLGTFAVSYIANKVLTQVTDNALSKRNEKWDVVKYIIVTNLAK 720
Db 661 AVILLEIPEIAPVLGTFAVSYIANKVLTQVTDNALSKRNEKWDVVKYIIVTNLAK 720
Qy 721 VNTQIDILIRKQKBALENQAEATKAIINYQVNYTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKQKBALENQAEATKAIINYQVNYTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKELNQCYSYLKNSMIPYGVKLEDPDASLKDALLKIYDNRGTILGOVDRKDK 840
Db 781 MININKELNQCYSYLKNSMIPYGVKLEDPDASLKDALLKIYDNRGTILGOVDRKDK 840
Qy 841 VNNLTSTDIPEQLSKYVDNORLLSTFTYEIKNIINTSILNLRVBSNHLIDLSRYASKINI 900
Db 841 VNNLTSTDIPEQLSKYVDNORLLSTFTYEIKNIINTSILNLRVBSNHLIDLSRYASKINI 900
Qy 901 GSKVNFDPIDKQNLQFLNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNISLNN 960
Db 901 GSKVNFDPIDKQNLQFLNLESSKIEVLKNAIVNSMYENFSTFWIRIPKYFNISLNN 960
Qy 961 EYTTINCENNSGKWSLNGEIIITWLDQTEIKQVVKYSQMINISDYNIRWIFVTIT 1020
Db 961 EYTTINCENNSGKWSLNGEIIITWLDQTEIKQVVKYSQMINISDYNIRWIFVTIT 1020
Qy 1021 NNRNNKSIYNGRLIDQKPTISNIGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRNNKSIYNGRLIDQKPTISNIGNTHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKOLDYQNSGIIKDFWGDYLOYDKPYMLNLYDPNRYVDNNVGIRGYMLKGP 1140

Db 1081 EKEIKOLDYQNSGIIKDFWGDYLOYDKPYMLNLYDPNRYVDNNVGIRGYMLKGP 1140
Qy 1141 GSVMTTNIYLNLSLYRGTKFIKKYASGNKDNIVRRNDRVYINVVKNEYRLATNASQA 1200
Db 1141 GSVMTTNIYLNLSLYRGTKFIKKYASGNKDNIVRRNDRVYINVVKNEYRLATNASQA 1200
Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKVDQGITNKCXNKQDNNGNDIGFTGFHGFNNIAK 1260
Db 1201 GVEKILSALEIPDVGNLSQVVMKSKVDQGITNKCXNKQDNNGNDIGFTGFHGFNNIAK 1260
Qy 1261 LVASNNYNRQIERSSRTLCSSWEFIPVDDGWERPL 1296
Db 1261 LVASNNYNRQIERSSRTLCSSWEFIPVDDGWERPL 1296
RESULT 3
S33411
botulinum neurotoxin type F - Clostridium baratii
C;Species: Clostridium baratii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: UNIPROT:Q45851; EMBL:X68262; NID:g49138; PID:CAA48329.1; PID:g49138
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
Query Match 36.7%; Score 2504.5; DB 2; Length 1268;
Best Local Similarity 41.4%; Pred. No. 1.1e-101; Indels 107; Gaps 36;
Matches 552; Conservative 242; Mismatches 433;
Qy 1 MPFVNKQFNKDPVNGVDIAYIKIPNAGOMQPVKAFKIHKNKIWIIPERDFTTPEBGDLN 59
Db 1 MPVNNFNVDPIINNTTILYMKMPYVEDSNKYKAFKIHKNKIWIIPERN-IIGKKPSDF 59
Qy 60 NPPEAKQVPVSYDYDSTYLSKDNEKNYLKGVTKLPERIYSTDLGRMLLTISVIRGIPFWG 119
Db 60 YPPIISLDSSGSSAYDPEYNTLTDKDRFLKTVIKLFNRINSNPAGQVLLBEIKNGKPYLG 119
Qy 120 GSTIDTELKVIDTNCIN-----VIQPDGYSYRSEBELNLIIGPSADIIQFECKSFGEV 172
Db 120 -----NDHTAVNEFCANNRSTSVETIKESNGTTDSMLNLIIVILGPGNII--ECSTFPVRI 172
Qy 173 L-----NLTRNGYSGTQYIRFSPDFTGFPESLEVDNPLLGAGKATDPAVTLAHELI 226
Db 173 FPNNAIYDSEKFGSGIQLMSFSTEYEAENDTDL-----FIADPAISLAHELI 222
Qy 227 YAGHRLYGI-AINPNRVKNTNAYEMSGLEVSFEELRTFGHDAKPIDSLQENEFRLY 285
Db 223 HVLHGLYGAGVTKVKKVIEVDQCALM-AAEKDIKIEBIFTFGGQDLNLIITNSTNQKIYVI 281
Qy 286 YNNKFKDIASTL---NKAKSIVGTTASLQYMKVNFKEKYLLEDTSKGFSVDKLFKDKLY 342
Db 282 LLSNYTAIASRUSQVNRNNSALNTT---YKKNFFQMKYGLDQDSNGNYTVNISKFNAYI 337
Qy 343 KMLTEIYTEDNFVKFVKVNLNRYLNFDAKVFKNITVPKVNITYYDGNLRLNLTNLAANFN 402
Db 338 KKLFS-FTCEDLAQKQVKNRNYLPHFKPFRLLDLDLDDNIIYSISEGFI--GSLRVNN 394
Qy 403 GQNTENNNMTFKLNFTGLFEFYKLCVRGIIITSK-TKSLDKGYNKALNDLCIKVNNWD 461
Db 395 GQINILNRSIVGPIPD-NGLVERFVGLC-KSIVSKGKTK-----NSLCTIKVNNRD 442
Qy 462 LFPSPEDNFT-NDLNGEBEITSDTNIEAAEENISLDLIQOYLYTTFNFDNEPENISLENL 520
Db 443 LFPVASESSYNGINGINSPKEI-DDTTITNNNYKKNLD---EVLIDYNSD-----AIPNL 492

Qy 521 SSDIIGQL----ELMPNIEFPNG----KKYELDKYTMFHYLRAOBEPHECKSRIALTNSV 572
Db 493 SSRLLNTAQNDSDSPKYD--SNGTSIKBYTQDLKLVNFFLYLAQKAPAGESAISLTSSV 550
Qy 573 NEALLNPSRVVTPSSDYVKKVKATAAAMFLGVEQVLVDFDTESEVSTTDKADIITI 632
Db 551 NTALLDASKYTFSSDFINTVKNPQVQAALFISWQQVINDFTTEATQKSTIDKIADISL 610
Qy 633 IIPYIGPALNIGMMLYKDDFVGALIFSGAVILLEFFIEPIAIPVLGTALVSYI-----AN 687
Db 611 IVPYVGLALNIGNEVQGNFKPEATELLCAGILLFEPPELLIPTILVPTIKSFINSDDSKN 670
Qy 688 KVLTVQIDNALSKNEKQWVYKVIYTNWLAKVNTQIDILRKQKKEALENQAEATKAI 747
Db 671 KI--IKAINNALRELKWKVEYSWISVSNMLTRINTQFNKRKEQMYQALQVQDVGKII 728
Qy 748 NYQYNQVTEEEKNI--NFNIDDLSSKLNESINKAMINIKFLNQCVSVYLMNSMPIYGV 805
Db 729 EYKNNYTLDEKRLRAEYNIYSIKEELNKKVSLAMQIDRFLTESSISYLMKLINEAKI 788
Qy 806 KRLDFDASLKDALKYIYDNRGTL-IGQVDRLLKDKVNNLTSTDPQLSKYVDNQRLLS 864
Db 789 NKLSEYDKRVQYLLNYLENSSTLGTSSVPELNNLVSNLTNLSIPPSELYTNDKTLIH 848
Qy 865 TFFEYIKNINTSLNLRYESNHLIDLSRVASKINIGSKVNFDPIDKNQIOLFLESKI 924
Db 849 ILIRFYKRIIDSSILNKKIENRRFIDSGYGSNISNGDIYIYSTNRNQFGISRSRSEV 908
Qy 925 EVILKNAIVNMSYENFSTFWIRIPKYPNISLSNNEYTIINCM-ENNSGKVKSLNYGEI 983
Db 909 NITQNTIYNSRYQNFVSFWRIPKYNLKNLNEYTIINCMNRNNSGKWSLNYNNI 968
Qy 984 IWTLODQIEIKQVVKYSQMINISDYINRWIFVTITNRLNLSKIYNGELIDOKELISN 1043
Db 969 IWTLODQTTGNQKLVNFYTMIDISDYINKWTFVTITNRLGHSKLYNGNLTDQKSLN 1028
Qy 1044 LGNTHASNNIMFKDGGCDTHRYIWIYFNLFDKELNEKEIKOLYDNQSNIGILKDFWGD 1103
Db 1029 LGNIHVDNIIIFKIVGCDNT-RYVGIRYFKIFNMELDKTELETLYHSEPDSTILKDFWGN 1087
Qy 1104 YLOVDKPYMLNLDYDPNKYVDVNNVINGIRGYMYLKGPRGSVMWTNYLNSLYRGTPIIK 1163
Db 1088 YLLYNNKYYLENLLKPNMSVTKNS-----DILNINRQGIYSKTIWIFSNARLYTGEVIR 1143
Qy 1164 KYAS---CNKNNIVRNDRVYINVVKNKEYRLATNSAQGVKILSALEIPDVG-NLSQ 1219
Db 1144 KVGSTDTSTNDFVRKNDTVYINVDGNSYQLYADVSTSAVEKTIKLRISNSYNSNQ 1203
Qy 1220 VVWKSNDQGITNKKCNLQDNGNDIGFIGHQFNIAKLVASNWTNRQIERSRSLTG 1279
Db 1204 MIIDS-----IGDCTWNFKTNNGNDIGLGFH-LNN--LVASSWYKYNRNTRNG 1254
Qy 1280 CSWEPFVDDGNGE 1293
Db 1255 CFWSFISKEHGWOE 1268

RESULT 4
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 09-Jul-2004
C:Accession: S21178; S48107; JH0257; B35294; A60027; S1811
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A:Reference number: S21178; MUID:92174922; PMID:1541280
A:Accession: S21178
A:Molecule type: DNA
A:Residues: 1-1252 <WH>
A:Cross-references: UNIPROT:Q00496; UNIPROT:Q45862; EMBL:X62683; NID:940397; PIDN:CAA445
R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A:Reference number: S48103; MUID:94013372; PMID:8408542
A:Accession: S48107
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 616-982 <CAM>
A:Cross-references: EMBL:X70815; NID:9407786; PIDN:CAA50146.1; PID:9407787
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type
A:Reference number: JH0256; MUID:92181428; PMID:1543481
A:Accession: JH0257
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-176, 'R', 178-197, 'C', 199-339, 'R', 341-772, 'I', 774-962, 'FE', 965-966, 'R', 968-1
A:Cross-references: EMBL:X62089; NID:940393; PIDN:CAA43999.1; PID:940394
A:Experimental source: strain Beluga
R:Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A:Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A:Reference number: A3294; MUID:90264400; PMID:2160960
A:Accession: B35294
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-176, 'R', 178-252 <BIN>
A:Experimental source: strain Beluga
R:Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A:Reference number: A60027; MUID:90344918; PMID:2116911
A:Accession: A60027
A:Molecule type: protein
A:Residues: 420-427 <GIM>
A:Experimental source: strain Beluga
A:Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C:Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
F:1-427/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F:423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F:412-426/Disulfide bonds: #status predicted
Query Match 34.3%; Score 2339; DB 2; Length 1252;
Best Local Similarity 39.9%; Pred. No. 1.8e-94;
Matches 535; Conservative 249; Mismatches 421; Indels 136; Gaps 43;
Qy 1 MPFVNQFNKDPVNGVDIAYIKIPNAGQNPVKAFKHNKIWIPIERDTF-TNPEEGDL 59
Db 1 MPKIN-SFYNDPVDNRITLYIK--PGGCOEFYKSFNIMKNIIWIPERNVIGTTPQ--DF 55
Qy 60 NPPEAKQVPVSYVDSTLTDNEKNYKLVTKLPERIYSTDLGRMLTSIVRGIPFWG 119
Db 56 HPPTSLKNGDSSYDYPNYLOSDEKRFKIVTKIFNRIINNLSGGILLSELSKANPYLG 115
Qy 120 G-STIDTELKVIDTNCINVIQPDGSYRSEELNVIIGPSADIIOPECKSGHEVLN---L 175
Db 116 NDNTPDNQPHIGDASAVEIKFNSGSDILLPNVIIMGAEPLD--FETNSSNISLRNNYP 173
Qy 176 TRNGYGTQVIRSPDPTFGFEESLEVDVTNPLLCAGKFAFDPAVTLAHLIYAGHRLY-- 233
Db 174 SNHGFSGIAIVTSPSEYSPFNNDN-----SMNEFIQDPALTLHMLHLSLHGLYGA 224
Qy 234 -GIAINENRVFKVN---TNAYEMSGLEVSFEELRTFGGHDAKFDISLOENFELYNNK 289
Db 225 KGITTKYIQQKNPLTN-----IRGINI--EEFLTGGTDLNIIITSAQNDIVTNLLAD 278
Qy 290 FKDIASLTNKAISVGTGTTASLQYMNKVFKEKYLSEDTSGKFSVDKLFKPKLYKMLTEIY 349
Db 279 YKIKASKLSKQV---SNPLNPNKYDVEAKYGLDKDASGIYSVYNKFNKFNDFKGLYS-F 334

Qy 573 NEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWEOVLVYDFTDETSEVSTTDKIADITI 632
 Db 542 DTALLEOPKITYTFSSSEFINNKNKPVQALFVGNIQOVLVDFTEANQKSTVDKIADISI 601
 Qy 633 IIPYIGPALNIGNMLKXDDPVGALIFSGAVILLEFPEIAIPVLGTFALVSYIA---NKV 689
 Db 602 VVPYIGLALNIGNEAOKGNFKDALELIGAGILLEFPELLIPTILVFTIKSFLGSSDNKN 661
 Qy 690 LTQVOTINALSKNEKWDVYKIVTWNLAQVNTQDILIRKKKEALENOAETKALINY 749
 Db 662 KVIAINNALKEDKKEVYSFVSNWMTKINTQFNKRKEQMYQALQONVAKALIES 721
 Qy 750 QYNOYTEEEKNNI--NPNIDLLSKLINESINKAMINIKFNLCQSVSYLMSMIPYGVKR 807
 Db 722 KYNSYULEEKNELTNKYDIEQIENELNOKVSIAMNIDRLTSSISYLMKLINEVKINK 781
 Qy 808 LEDPDLAKDALLKYIYDNRGTIGQVDR-LKQVNNLTSTDPQLSKYVDNORLLSTF 866
 Db 782 LREYDENVKTYLDYII-KHGSILGESQBELNSMVIDTLANNSIPFKLSSYTDKILISYF 840
 Qy 867 TEYIKNINTSIILNRYESNHLDLRYASKINIGSKVNFDPIDKNOIQLFNLESSKIEV 926
 Db 841 NKFKRIKSSSVLMRYKNDKYDVTSGYDSNININGVYKYPTKNQFGIYNDKLSSEVNI 900
 Qy 927 ILKNAIVNSMYENFSTFWIRIPKYPNSI-SLANNEYTIINCM-ENNSGKVSINYEII 984
 Db 901 SONDYIIYDNYKYNKFSIFWVRPNYDNKIVVNNVEYTIINCMEDNNSGKVSINHNEII 960
 Qy 985 WTLQDQEIQRVVKYKYSQMINISDYINRWIFVTITNNLNNSKIYINGRLIDOKPISNL 1044
 Db 961 WTLQDNGINGKLAFAFYNGANGISDYINKWIFVTITNDRLGDSKLYINGRLIDOKSILNL 1020
 Qy 1045 GNTHASNNIMFKDGCRTTHRYTWIKYFNLFDEKELNEKEIKOLYDQNSGILKDFWGDY 1104
 Db 1021 GNTHVSDNILFKVNCYSY-RIYIRYFNIPDKELDETEQTLNNPENNANILKDFWGY 1079
 Qy 1105 LOYDKPYMLNLYDPNKYVD-----VNVGIRGYMYLKGPRGSVMYMTNIVLNSLYR 1156
 Db 1080 LLYDKEYLLNVLKPNFNRRTDSTLSINNI-----RSTILLANRLYS 1123
 Qy 1157 GTKFIILKY-ASGNKDNIVNRNDVYINVVVKNKEYRLATNASOAGVEKILSALEIPDVG 1215
 Db 1124 GIKVKIORVNNSTNDLVRKNDQVYNFVA-SKTHLLPLYADTATTNK-EKTIKSSSG 1181
 Qy 1216 N-LSQVVMKSKDQGITKCKMLODNGNDIGFHFQFNNTAKLVASWVYNQIERS 1274
 Db 1182 NRFQVVMVMS-----VGNCTMFKNNNGNIGLGF-----KADTVVASTWYTHWRDN 1231
 Qy 1275 SRTLGCSWEFIPDDGWGER 1294
 Db 1232 TNSNGFFWNFISEHGWQEK 1251

RESULT 6
 I40813
 neurotoxin type F - Clostridium botulinum
 C:Species: Clostridium botulinum
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
 R:East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
 C:Accession: I40813; S48108
 A:Residues: 1-1274 <RES>
 R:Campbell, K.D.; Collins, M.D.; East, A.K.
 J. Clin. Microbiol. 31, 2255-2262, 1993
 A:Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
 A:Reference number: 140644
 A:Accession: I40813
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <RES>
 A:Cross-references: UNIPROT:P30996; GB:M92906; NID:g144866; PID:AAA23263.1; PID:g144867
 A:Title: Gene probes for identification of the botulin neurotoxin gene and specific id

A:Accession: S48108
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 634-1002 <CAM>
 A:Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
 C:Superfamily: tetanus toxin
 C:Keywords: neurotoxin

Query Match 33.7%; Score 2296.5; DB 2; Length 1274;
 Best Local Similarity 39.4%; Pred. No. 1.3e-92;
 Matches 523; Conservative 253; Mismatches 465; Indels 85; Gaps 39;

Qy 1 MPVKNQFNYKDPVNGVDIAYIKIPNAGQMOP-VKAFKIHNNKIWIPIPERDTP-TNPEEGD 58
 Db 1 MPVAINSFNVPNDTILYMQIPIYEEKSKYKAFEMRNVMYIIPERTITGTNP--SD 58
 Qy 59 LNPPEAKQVPVSYDSTYLDNEDKNYLKGVTKLFRYISTDLGRMLLTSTVRGIPFW 118
 Db 59 FDPASLKGNGSSAYDPNVLTTDAEKDRYLKTKIKLFRKINSNPAGKVLLOEISYAKPYL 118
 Qy 119 GG--STIDTELKVIDTNCINVIQPDGYSRSEELNLVIIIGPSADIIQFEC----KSGHEV 172
 Db 119 GNDHTIDFESPVTRITTSVNIKLSNVESMLNLVLGAGDPDIPESCCVPVRKLLDPDV 178
 Qy 173 LNLTRN-GYGSQYIRPSPDFTGPEESLEVDTNPLLGAGKATDPDPAVTLAHLIYAGHR 231
 Db 179 VDPDSNYGSGINIVTFSPSEYETFN---DISGHNSSSTESFTADPAISLAHELIIHALHG 235
 Qy 232 LVGI-AINPNRVKNTNAYEMSGLEVSEELRTFGGHDAKFIDSLQENHEPLYYNKF 290
 Db 236 LYGARGVTEETEVK-QAPLMAETKPIRLEELFTGGQDLNITTSAMKEKINLLANY 294
 Qy 291 KDIASINAKSIVGTFTASLOY----MKNVFKEKYLLEDSTGKFSVDKLFKDKLYKMLT 346
 Db 295 EKIAIRLSEVNS-----APPEVDINEYKDYFQWKYGLDKNADGYSYTVNENKFEIYKLY 349
 Qy 347 EITENDFVKFVKLNARKTYLNFDAVF-KI-NIVFKVNTIYDGNLRTNLAANFNGQ 404
 Db 350 S-FTESDLANKFKVKCRNTY--PIKYEFUKVPLNLLDDDIYTVSEGENI--GNLAVNRGQ 404
 Qy 405 NTEINNMFTKLNFTGLFPEFYKLLCVGRIITTSKTSKLDKGNKALNDLCIKVNNMDLFF 464
 Db 405 SIKLNPKIIDSIPD-KGLVEKIVKFC-KSVIPRKG-----TKAPPLCIRVNSSELFP 455
 Qy 465 SPSEDNFT-NDLNKGBEITSNTIEAAENISLDLIQOYLYTFNFDNENIENIEMSSD 523
 Db 456 VASESSYNENDINTPKEDIDTTLNANNYRN-NLD---EVILDYNSQTIPO-ISRNLNT- 509
 Qy 524 IICQLELMPNIERFPNG---KKVELDKYTMHRYLRAQEPHEGKSRIALTNSVNEALLNP 579
 Db 510 LVQDNSYVPRYD--SNGTSEIEYDVVDVFNFFYLHAQKVPGETEINISLTSSIDTALLEE 567
 Qy 580 SRVYTFSSDYVKKVKNKATEAAMFLGWEOVLVYDFTDETSEVSTTDKIADITIIPYIGP 639
 Db 568 SK-DIFPSSSEFIDTINKPVNAALFIDWISKVIRDTEATQKSTVDKIADISLIVPVGL 626
 Qy 640 ALNIGNMLYKDDPVGALIFSGAVILLEFPEIAIPVLGTFALVSYI---ANKVLTVOTID 696
 Db 627 ALNIIIEAEGNFEAEFELGVGILFVPELTPVILVFTIKSYIDSYENKKAINKAIN 686
 Qy 697 NALSCKNEKWDVYKIVTWNLAQVNTQDILIRKKKEALENOAETKALINTQYNOYTE 756
 Db 687 NSIIEAKKWEKYSIVTWSNWLTRINTQFNKRKEQMYQALQONVDAIKTAIEKYNNTS 746
 Qy 757 EEKNNI--NPNIDLLSKLINESINKAMINIKFNLCQSVSYLMSMIPYGVKLEDPDAS 814
 Db 747 DEKNRLESEYNNINIEELNKKVSLAMKNIERFWTESSISYLMKLINEAKVGLKKYDNH 806
 Qy 815 LKQALLKYIYDNRGTIGQVDR-LKQVNNLTSTDPQLSKYVDNORLLSTFTEYIKNII 874
 Db 807 VKSDLLNYLDHRSILGEQTNELSDLVSTLNSSIPFELSSTYNDKILITVFNRLYKKIK 866
 Qy 875 NTSILNRYESNHLIDLRYASKINIGSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVY 934

Db 867 DSSILDMRYENKFPIDISGYSNISINGNYIYSTNRQFGIYNSRLSEVNIQAQNNDIY 926
Qy 935 NSMYENFSTFWIRIPKYFNSISLNNYTIINCM-ENNSGWKYSLN-GEIITWLODT 990
Db 927 NSRYQFISFWIRIPKHYKPMHNRREYTIINCMGNNNSGWKISLRVDRCEIITWLODT 986
Qy 991 QETKQRVVKYQOMINSDYINRWIFVTITNNRLNNSKIYINGRLIDOKPISLGNTHAS 1050
Db 987 SGNKENLIPRYELNRIISVINKWIFVTITNNRLNNSKIYINGRLIIVKESISLNDIHVS 1046
Qy 1051 NNIMFKLDGCDTHRYIWKYFNLFDKELNEKEIKDLYDNQNSGILKDFWGYLOYDKP 1110
Db 1047 DNLIFKIVGC-DETYGIRYFKVFNTDELTKETIETLYSENPSPILKNYWGNYLLYK 1105
Qy 1111 YMLNLDPKYYDVNVNGIRGYMYLKGPRGSVMTNIVLNSLSYRGTETIKKYAS-- 1167
Db 1106 YLFLNLRKQYITLNS---GILNINQORG-VTEGVSFLNYKLYEGVEIIRKNGPIDI 1160
Qy 1168 GNKDNIVRNRDRYINVVVKNKEVRLATNASQAGVEKILSALEIPDVGNLSQVVMKSKN 1227
Db 1161 SNTDNFVRKNDLAYINVDVGRVRYLAD-TKSEKEKIIRTSNLND--SLGQIIVMDS-- 1215
Qy 1228 DOGITNKKVNLQDNGNDIGFIFGHPFNIAKLIVASWYNRQIERSRILGCSWEPIPV 1287
Db 1216 ---IGNNCTMNFQNNNGSLGLGFHSNN---LVASSWYNNIRRTSSNGCFWSSISK 1268
Qy 1288 DCGWGE 1293
Db 1269 ENGWKE 1274

RESULT 7
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 33.5%; Score 2285; DB 2; Length 1297;
Best Local Similarity 39.6%; Pred. No. 4.2e-92;
Matches 531; Conservative 217; Mismatches 502; Indels 90; Gaps 29;

Qy 1 MPFVKQFNKYKDPVNGVDIAYIK-IPNAGQMPKAFKHNKIWIPIERDTF-TNPEEG 58
Db 1 MPVNIKNFNNDPNNDDIIMWEPFDPGPGTYKAFRIIDRIWIPIERDTFPGPDQFN 60
Qy 59 LNPPEAKQVPVSYDSTYLSTDNKDNLYKGVTKLPERIYSTDLGRLMALTSLVRGIPFW 118
Db 61 ASTGVFSKDV-YEYDPTLYTKDAEDQKFLTKMFLNRSKPSGORLLDMIVDAIPYL 119
Qy 119 GGSTIDTELKVIDTNCINLV-----IQPDGSYRSEEL--NLVIGPS---ADIIQPECKS 167
Db 120 GNAS--TPPKFAANVANVINKKIIQPGADQIKGLMTWLLIIFGPGVLSNFTDSMIM 177
Qy 168 FGHEVLNLRNGYSGTQYIRFSPDFTFGFESLEVDTPNLLGAGKATDPAPVLAHELIIY 227
Db 178 NGHSPIS--EGFGARMIRFCPSCLNVNQNENKDSIFSRRAYFADPALTLMEHLIIH 234
Qy 228 AGHRLYGIANPNRVKNTNAYEMSGLEVSFEELTFGCHDAKFIDSLQENEFRIYY 287
Db 235 VLHGLYGIKIS-NLPITPTNTEKFFMQHSDFPVQAEELYTFGCHDFSVISPSD-----MNIY 289

Qy 288 NK----FKDIATSLNKAKSIVGTASLQYMNKVNFKKYLLSEDTSGKFSVDKLFKLYK 343
Db 290 NKALQNFODIANRLNIVSSAQSGSIDISLYKQIYKKNYDFVEDPNGKYSVDKDFKLYK 349
Qy 344 MLTEIYEDNFVKFPFKVLARKTYLNFDPKAVFKINIVPKVNYTYI---DGNLRLNTNLAAN 400
Db 350 ALMFGTETTLAGEYGIKTRYSY--FSEYLPPIKTEKLDNTIYTQNEGENIASKNLUKE 407
Qy 401 FNCQNTIEINNMFTKLKNFTGLPEFVKILCVRGIIITSKTSKSLDKGVNKALNDLICIKVNNW 460
Db 408 FNCQKAVNKEAYEES--LEHLIVIRIAMCKPMYK-----NTGSEQCIIIVNE 456
Qy 461 DLFPSPSEDNFTNDLNGBEISDNTWIEAABENISLD-LIQYLYLTFNFONEBENISIEIN 519
Db 457 DLFFIANKDSFSKDLAKAETIAYNTQNTTENNFSIDQLILDNDLSSGIDPLNENTPEFT 516
Qy 520 LSSDI-----IGOLELMPNIERFPNGKKYELDKYTMFHYLRAQEFPHGKRIALTNSVNE 574
Db 517 NFDDIDI PVYIKOSAL-----KKIPVDGDSLFYELHAQTFPSPNIENLQJLNSLND 566
Qy 575 ALNPNRGRVYTFSSDYVKKVKNKATEAAMFLGWGEVLVYDFDTSETSEVSTTDKIADIRIII 634
Db 567 ALRNNKVTYTFSTNLVEKANTVVGSLFVNNVKGVIDDFESTQKSTIDKYSDVSIII 626
Qy 635 PYIGPALNIGNMLYKDDFVGALIFSGAVILLEPIEPIAIPVLGTFPALVSVYANKVLTVQT 694
Db 627 PYIGPALNVGNETAKENFKNAPEIGGAAILMEFIPELIVPIVGFTLESYVGNKGHIIMT 686
Qy 695 IDNALSKRNEKDEVYKIYVNWLAQNTQIDILRKKMKEALENAQEAATKALINYQNOY 754
Db 687 ISNALKKRQKQWTDMYGLIVSQWLSTVNTQFYIKERMYNALNNQQAIEKIIDQYNR 746
Qy 755 TEEKNNINFDISSLKLNESINKAMININKELNOCVSYSYLMNSMIPYGVKLEDFDAS 814
Db 747 SEEDKNNINFDNDIDFKLNQSLNLAINDIDFQNCISYLMNRMIPYAVKLLKDFDND 806
Qy 815 LKDALIKYIDNRGTILIGQVDRLLDKVNNTLSTDIPFQLSKYVDNQRLLSFTFEYIKNII 874
Db 807 LKRDLLEYIDTNELYLLEWNLKSVNRHLKDSIPFDLSLYTKDTILIQVFNYSNIS 866
Qy 875 NTSILNLRVSNHLIDLRYAKINISGVKPNFDPIDKQIOLFNLSKKEVILKNAIVY 934
Db 867 SNAILSLSYRGRLIDSSGYGATMVGSDVIFNDIGNQPKLNNSENSTIAHQSKFVY 926
Qy 935 NSMYENFSTFWIRIPKYFNS---ISLNNEYTIINCMNNSGWKYSLNIGEIIWLODTQ 991
Db 927 DSWFDFNSINFWRTPKYNNDIQTLYQNEYTIISCIKNDSGWKVSIKGNRIIWLIDVN 986
Qy 992 EIKQRVVKYQOMINSDYINRWIFVTITNNRLNNSKIYINGRLIDOKPISLGNTHAS 1051
Db 987 AKSKSIFFEYSIKDNISDYINKWFSITITNDRLGANAIYINGSLKSEKILNDRINSSN 1046
Qy 1052 NIMFKLDGCDTHRYIWKYFNLFDKELNEKEIKDLYDNQNSGILKDFWGYLOYDKPY 1111
Db 1047 DIDFKLINCTDTTKFWIKDFNIFGBELNATEVSSLYWISQSTNTLKDFWGNPLRYDTQY 1106
Qy 1112 YMLNLDPKYYDVNVNGIRGYMYLKGPRGSVMTNIVLNSLSYRGTETIKKYASG-- 1168
Db 1107 YLFNQGMQNIYIKFS---KASMGETAPRTNF--NNAAINYQNLGLRLIIFKASNRN 1161
Qy 1169 -NKDNIVRNRDRYINV-VYKKEVRLATNASQAGVEKILSALEIPDVGNLSQVVMKSK 1226
Db 1162 INNDNIVREGDIYILNIDNISDESRYVVLVNSKEIQTLFLAPINDPFTFYDLQIK-K 1220
Qy 1227 NDOGITNKKVNLQDNGNDIGFIFGHPF-----NNIAKLIVASWYNRQIER-- 1273
Db 1221 YYEKTTYNCQI-LCEKDTKTFGLFGICKFKYQYGYVMDTYDNY--FCISQWYLRRISENI 1277
Qy 1274 SSRTLGCSWEFIPDDGWGE 1293
Db 1278 NKURLGCMQWFIPIVDSGWTE 1297

RESULT 8

A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N/Alternate names: botulinum neurotoxin type B (BoNT/B)
C/Species: Clostridium botulinum
C/Date: 19-Dec-1993 #sequence revision 18-Nov-1994 #text change 09-Jul-2004
C/Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R/Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 59, 2345-2354, 1992
A/Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type B neurotoxin
A/Reference number: A48940; MUID:92384550; PMID:1514783
A/Accession: A48940
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1291 <WHE>
A/Cross-references: UNIPROT:P10844; GB:M81186; NID:G144734; PIDN:AAA23211.1; PID:G144735
A/Experimental source: type B, Danish
A/Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publication
R/Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A/Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A/Reference number: S48105; MUID:9403372; PMID:8408542
A/Accession: S48105
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 634-994 <CAM>
A/Cross-references: EMBL:X70817; NID:G407782; PIDN:CAA50148.1; PID:G407783
A/Experimental source: proteolytic type B, strain NCTC 7273
R/Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A/Description: Partial amino acid sequence of botulinum neurotoxin type B and comparison
A/Reference number: S21575
A/Accession: S21575
A/Molecule type: DNA
A/Residues: 36-217, 'G', 219-224, 'S', 226-246 <SZA>
A/Cross-references: EMBL:Z11934; NID:G40383; PIDN:CAA77991.1; PID:G40384
R/Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K.
J. Biol. Chem. 267, 14721-14729, 1992
A/Title: Minimal essential domains specifying toxicity of the light chains of tetanus toxin
A/Reference number: A42871; MUID:92340509; PMID:1634516
A/Accession: A42871
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-313, 'S', 315-451 <KUR>
A/Experimental source: strain Okra
A/Note: sequence extracted from NCBI backbone (NCBIP:109365)
R/DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A/Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A/Reference number: S07155; MUID:89000987; PMID:3139097
A/Accession: S07155
A/Molecule type: protein
A/Residues: 2-29, 'M', 31-45 <DAS>
A/Accession: S08562
A/Molecule type: protein
A/Residues: 442-463, 'R', 465-467 <DA2>
R/Schmidt, J.J.; Sathyanarthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A/Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A/Reference number: S07128; MUID:85197963; PMID:3888113
A/Accession: S07128
A/Status: preliminary
A/Molecule type: protein
A/Residues: 2-16 <SCH1>
A/Accession: S08573
A/Status: preliminary
A/Molecule type: protein
A/Residues: 2-17 <SCH2>
A/Accession: S08574
A/Status: preliminary
A/Molecule type: protein
A/Residues: 442-459 <SCH3>

R/Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A/Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic
A/Reference number: S27125; MUID:93063293; PMID:1331807
C/Contents: annotation
C/Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses
C/Genetics:
A/Genes: bont/b
C/Function:
A/Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C/Superfamily: tetanus toxin
C/Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F/2-441/Product: bontoxilysin B light chain #status experimental <LIGHT>
F/442-1291/Product: bontoxilysin B heavy chain #status experimental <HVT>
F/230-234/Binding site: zinc (His) #status predicted
F/231/Active site: Glu #status predicted

Query Match	33.3%	Score	2271;	DB 1;	Length	1291;	
Best Local Similarity	38.7%	Pred. No.	1.7e-91;				
Matches	523;	Conservative	234;	Mismatches	477;	Gaps	27;

Qy	1	MPFVNQENYKQDVGVGDVIAIKIPNA-GQMOPVKAFKHNIKIWIPIERDTF-TNPREGD	58
Db	1	MPVTINNFNDPIDNNIIMMEPPFARGTGRYYKAFKIDRIWIPIERTFYGKEDFN	60
Qy	59	LNPPEAKQVPVSVDSTYLSTDNEDKNVYKLVTKLFIERYSTDLGRMLTTSIVRGIPFW	118
Db	61	KSSGIFNRDV-CREYDPDYLTNDKNIFLQTKLFNRKSPKSGKLEMIINGIPYL	119
Qy	119	GGSTIDTELKVIDTNCINV-----IOPDGSYRSE---LNLVIIGSADIIOECSFGH	170
Db	120	GDRRV--PLEEFNTIASVTNKLINSPGVERKGFANLIIFGPGVLNENETDIDI	177
Qy	171	EVLNLTNRGVGSGTYIRFSPDFTFGFEESLEVDNPLLGAGKATDPAPVLAHELIYAGH	230
Db	178	QNHFAREGGGMQMKFCPEYVSFVNNQENKASIFNRRGFSDDALLMHELHVLH	237
Qy	231	RLYGIAT-----NPNRVFKVNTNAYEMSGLEVSEELRTFGGHDAKFDLSQENFR	283
Db	238	GLYGIKVDLLPIVPNEKKFMQSTDA-----IQAEELYTFGGQDPSIITPSTDKSIY	289
Qy	284	LYYNNFKDIASLTNKA-KSIVGTTASLQMKVNFVEKYLSSDTSKGFSVDKLPDKLY	342
Db	290	DKVLQNRGIVDRNLNKLVCISDPNININIKNFKDKYKFVEDSEKYSIDVSEFDKLY	349
Qy	343	KMLTEIYTDNFVKFPKVLNRKTYLNFDAVFKI-NLVPKVNTIYDGNLRTNLAAAF	401
Db	350	KSLMFGFTETNIAENYKIKTRASFSDSLPPVKIKMLLDNEIYTIIEGFINSKDMEKEY	409
Qy	402	NGQNTENNNTKLNFTGLFEFFYKLLCVRGHITSKTSKSLDKGYNKALNDLCIKVNNWD	461
Db	410	RGQNKAINKQAYEISK--EHLAVYKIOMCKSV-----KA-PGICIDVDNED	453
Qy	462	LPFSPEDNFTNDLNGEETSTNTNTEAAENISLDLIQOYLYTFNFDNEPENISLENLS	521
Db	454	LFPIADKNSFDOLSKNERIEYNT-----QSNYIENDF---PINELI--LD	494
Qy	522	SDIIGOLEL-----MPNIREFPNKGYELDKYTMFHYLRAQEFEGHKSRIA	567
Db	495	TDLSKIELPSENTESLTDENVDPVVEKOPAIKIFTDENTIFQYLYSTFPFLDIRDIS	554
Qy	568	LTSNVNEALNPSRVYTFSSDYVKVKVNAKATEAMFLGWVEQLVYDPTDTSVSTDKI	627
Db	555	LTSSFDALLFSNKVYSFFSMDVIKANKVVEAGLFAGVWKQIVDFVIEANKSNTMDKI	614
Qy	628	ADITIIPIYGPAINTGNMLYKODFVGALIFSGNAVILLEEIPETIAPVLGTALVSIVAN	687
Db	615	ADISLIVPYIGLALNVGNETAKGNFENAFIAGASILLFEPILLIPVVGAFLLSVYDN	674
Qy	688	KVLTVOTIDNALSKRNEKWDVYKVIYVNNLAKVNTQIDLIRKKMKALENQAEATKAI	747
Db	675	KNKIITIDNALFKRNEKSDMTGLIVAQWLSVTNTQFYTIKEGMYKALYQAALBEII	734

Db 1207 GSFIQPCNTFDYVYASQLFSSNATNRLGILSIGSYF----KLGDDYWFNHEYLIPIVI 1262
Qy 1271 -IERSRRL---GCSWEFIPVDD 1289
Db 1263 KIEHYASLLSTSTHWFVPASE 1285

RESULT 13
A49777
botulinum neurotoxin type C1 precursor - Clostridium botulinum (type C, strain c-st)
C:Species: Clostridium botulinum
C:Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: S11291; A35396; S22166; A49777
R:Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;
Nucleic Acids Res. 18, 4924, 1990
A:Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.
A:Reference number: S11291; MUID:90370487; PMID:2204031
A:Accession: S11291
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-84, 'P', 86-1291 <H>
A:Cross-references: UNIPROT:Q93HT3; EMBL:X53751; NID:gl4905; PID:CAA37780.1; PID:gl4906
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.;
Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990
A:Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxin
A:Reference number: A35396; MUID:91024998; PMID:2222445
A:Accession: A35396
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-669, 'R', 671-1291 <TS1>
R:Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of the gene for one of the components of hemagglutinin
A:Reference number: S22166
A:Accession: S22166
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1291 <TS2>
A:Cross-references: EMBL:X62389; NID:9558175; PIDN:CAA44263.1; PID:940390
R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.
Appl. Environ. Microbiol. 57, 1168-1172, 1991
A:Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and whc
A:Reference number: A49777; MUID:91282468; PMID:2059039
A:Accession: A49777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-607 <TS3>
A:Cross-references: GB:D90210
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 26.08; Score 1771; DB 2; Length 1291;
Best Local Similarity 32.88; Pred. No. 1.1e-69;
Matches 443; Conservative 244; Mismatches 542; Indels 122; Gaps 34;

Qy 1 MPFVKQFNQKDPVNGVDIAYIKIP-NAGQVQPVKAFKIHNKIWIPIERDTFTNPERGDL 59
Db 1 MPITINNPNYSDPVNKNILYDLTHLTLANEPEKAFITGNWIPDR--FSRNSNPNL 58

Qy 60 NPPEAKQVPVSYDSTYSLTDNEKNYKLVGTVKLPERIYSTDLGRMLLTSIVRGIPFWG 119
Db 59 NKPRVTSPPKSGYDNPYLSLSDSKDTFLKEIIKLFKRINSREIGBELIYRLSLTDIPFG 118

Qy 120 G-----STIDELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIQCECKSFCH 170
Db 119 NNNTPIETFD-----VDFNSVDVTRQGNWVKTGSINPSVITGPRENIIDPETSTF-- 173

Qy 171 EVNLNTRN-----GYGSGTQVIRSPDFTFGFEESLEVDNTNPLLGAKFA-----TDPVAV 219
Db 174 ---KLTNNTFAAQEGFALSIISIPRMLTYSNA-----TND-VGEGRFSKSEFCMDPIL 225

Qy 220 TLAHELIYAGHRLYGIAT-NPNRVFKVNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Db 226 ILMHHLNHAHNNLYGAIAPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGPTDIDLPKSA 285
Qy 279 ENEFRLLYYNKKFKDIASTLNKAKSIIVGTTASL-----QYMKNVFKEXYLLSEDTSGK 330
Db 286 RKYFEKALDYYSIAKLNSI-----TTANPSSFNKYIGEYKQKLIR-KYRFVWESSGE 339
Qy 331 FSVDKLKFKDKLYKMLTEIYTEDNFVFKFKVLNRTKTYLNFDFKAVFKINIVPKVNTIYVDF 390
Db 340 VTVNRNKFVELYNELTQITEFNYAKIYNVQNRKIYLSNVYTPVPTANILDDNVDYIQNGF 399
Qy 391 NLRNTWLAANFNGONTIENNMNFTKLKFTGLFEFVKLLCVRGIIITSKTSLOKGVNKAL 450
Db 400 NIPKSNLNLVFMQNLG--RNPALRKNVPENMLYLFK-FCHKAI---DGRSL---YNKTL 451
Qy 451 NDLCIKVKNWDLFPFSEONFTNDLKNGBEITSDTNIEAAEENISLDLIQYVLTNFNDN 510
Db 452 DCRELLVKNTDLPFIGDISDVKTDIRKDIINETEVIYYPDNVSD--QVILS---KN 505
Qy 511 EPNISIELNLSDDIIGOLELMPNIR-FPNKKYELDKYTMFHYLRAQEFHOKSRALT 569
Db 506 TSEHGOLDLLYPSIDSESEILPGENQVFDNRQNTQVDYLSYLYLESQKLSDNVEDFTFT 565
Qy 570 NSVNEALLNPSRVYTFESSDYVKKVKNKATEAAMFLGWVEOLVYDFETDETSEVSTDKIAD 629
Db 566 RSIEEALDNSAKVYTFPT-LANKVAGVQGGULFMANDVVEDFTNIIIRKDTLDKISD 624
Qy 630 ITIIPYIGPALNIGNMLKDDFVGALIFSGAVILLEFIPRIAPVLGTALSYIANKV 689
Db 625 VSAIIPYIGPALNISNSVRGNTEAPAVTGVTLLEAFPEFTIPALGAFVIYSKVQERN 684
Qy 690 LTVQTDNALSKNKKWDEYKIVTVTNLAKVNTQIDILRKMKKEALENOAETKALINY 749
Db 685 EIITKDNCLEQRIKWKDSYEMWMTGLSRIITQFNNTISYQYDLSINYQAGAKAKIDL 744
Qy 750 QYNQYTEEEKNNINFINDDLSSKLNESINKAMINIKFLNQCVSYLMNSMIPGVKRL 809
Db 745 EYKYSGSKENIKTSQVENLKNLSDVKISEAMNINKFIRECSVTYLFKMLPKVIDELN 804
Qy 810 DFDASLKALLKVIYONRGTLIQVDRDKQKVNNTLSTDIPFQLSKYVDNRQLLSTTEY 869
Db 805 EFDRTYAKLINLIDSHNILLVEVDKAKAVNNSQNTIPFNFISYTNNSLLKDIINEY 864
Qy 870 IKNIINTSLNRYENHLLDLSRYASKINIGSKVNFDPIDKNOIQLFNLESS-----KI 924
Db 865 FNNINDSKILQNRKNTLVDTSYNAEVSEEGDVLNPIFPD---FKLGSSGEDRGKV 921
Qy 925 EVILKNAIVVNSMYENFSTFWIRIPKYNFNSISLNNEYTIINCMEANSCHKVSLNYGEII 984
Db 922 IVTQENIIVVNSMYESFSISFWIRINKVWSNLP---GYTIIDSVKVNSGWSIGIISNVLV 978
Qy 985 WTLQDTQETIKORVVVKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL 1044
Db 979 FTLKQNEDEQSINFSYDINSNAPGY-NKQFFVTVNNMGNKNIYINGRLIDQKPISNL 1037
Qy 1045 GNTHASNINFKLDGCRDT-----HRYIWKYFNLPKELNEKEIKDYLNQNSGI 1096
Db 1038 TGINFSKTITFEINKIPDITGLITSDSDININWIRDFYIFAKELDGKIDINILFNSLQVNT 1097
Qy 1097 LKDFWGDYLDYKPYVMALYDPNKVVDVNNVIGIRGVWYMLKGPBGSVWTTNIYLNSSLR 1156
Db 1098 VKDYWGNDLRYNKYEYVWNIIDYLNRYWYANS-----RQIVFNTR-RNNDNFNE 1144
Qy 1157 GTFPIIKVAGSKNDKINVRNDRVYINNVVKNKEYRL-----ATNASQAGVEKILSA 1208
Db 1145 GYKILIKRIGTNTDTRVGRGDILYFDMTINNKAYNLFMKNETYADNHSTEDIYALGLR 1204
Qy 1209 LEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGNDIGTGFHQF-----NNI 1258
Db 1205 EQTKDINDNIIFQIQPMNNTYYVASQIFKSNFNGENISGICSGTFRFLGGDWYRHNYL 1264
Qy 1259 AKLVASWYNRQIERSRSLTGCSEWETIPVDD 1289

Db 1265 VPTVKGQNVASLLESTs-----THWGFVPVSE 1291

RESULT 14

S46431

botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)

N:Alternate names: BoNT/C1 protein

C:Species: Clostridium botulinum phage 1C

A:Variety: strain C 468

C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004

C:Accession: S46431; S49107

R:Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.

Mol. Gen. Genet. 243, 631-640, 1994

A:Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxic P

A:Reference number: S46426; MUID:94301293; PMID:8028579

A:Accession: S46431

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1291 <HAU>

A:Cross-references: UNIPROT:Q93HT3; EMBL:X72793; NID:G516171; PIDN:CAA51313.1; PID:G5161

A:Experimental source: strain C 468

C>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Superfamily: tetanus toxin

Query Match 26.0%; Score 1771; DB 2; Length 1291;

Best Local Similarity 32.8%; Pred. No. 1.1e-69;

Matches 443; Conservative 244; Mismatches 542; Indels 122; Gaps 34;

Qy 1 MPFVKNQFNYKDPVNGVDIAIKIP-NAGQOPKAPKIHKNKIWIPIERDTFTNPEGDL 59

Db 1 MPITINNFNYSDDPVNKNILYDLTHLNTLANEPEKAFRITGNWIPIPR--FSRNSNPNL 58

Qy 60 NPPEAKQVPVSYDSTVLSNDKONKLVGTVKLFERIVSTDLGRMLLTSIVRGIPFWG 119

Db 59 NKPRVTSKSGYDYPNLTSDSKOTFLKEIKLFRINSREIGELIYRLSTDIPFG 118

Qy 120 G-----STIDTELKVIDTNCINVIQPDGS--YRSEELN--LVIIGPSADIIOFECKSF 170

Db 119 NNNTPIITFD--VDVNSVDVKTROQNNVKTGINSPIVITGPENIIDPETSTF-- 173

Qy 171 EVMLNTRN-----GYSGTQVIRSPDFTGFPESLEVDNPLGACKEA-----TDP 219

Db 174 ---KLTNNNTFAAQGFAGLSIISIPRMLYSNA-----TND-VGEGRFSKSEFCMDPIL 225

Qy 220 TLAEHLIVAGRLXGIAI-NPNRVKVNNTNAYEMSGLEVSFEELRTFGGHDAKFIDSLQ 278

Db 226 LLMHNLHAMNLVGLIAPNDQTISSVTSNIFYQYVNVKLEYAIIYAFGGPTIDLI PKSA 285

Qy 279 ENFRLYYNKFDDIASTLNKAKSIVGTASL-----QYKNVFKKILLSEDTSGK 330

Db 286 RKYFEERKALDYVRSIAKRLNSI-----TTANPSSFNKYIGYKQKLIR-KYRFVVSSE 339

Qy 331 FSVKLFKDKLYKMLTEIYEDNFVKPKVLRNTYLNFDKAVPKINIVKVNITYIDGF 390

Db 340 VTNRNKFEVELYNELTQIFTEFNKIVYQNRKILYSNVVTPVTANILDDNVVDIQNGF 399

Qy 391 NLRNTLAANFNGQWTEINNNNFKLKNTGLFFYKLLCVRGHITSKRLSGYKAL 450

Db 400 NIPKSNLVLPWGQNLN-RNALRVNPNMMLYLFK-FCHKAI--DGRSL--YNTKL 451

Qy 451 NDLCKVNNWDLFPSPSEDNFTNDLKGEEITSDTNEAAEENISLDLIQYLYTFNPDN 510

Db 452 DCRELLVKNLTLPFTGDISDKVTDFLRKDINEETEVIYYPDNVSD--QVILS--KN 505

Qy 511 EPNISIEISDIIQGLKELMPNTER-FPNKGKVELDKYTMFHYLRAQEPHKGSRALT 569

Db 506 TSEHQDLQLLVPISDSESEILFGENQVYDNRQNVVDVNSYYLESQKLSNDVEDFTFT 565

Qy 570 NSVNEALLNPRVYFFSSDYKVKNKATEAAMFLGWVEQLVYDPTDTSVSTTDKID 629

Db 566 RSIEEALONSARKVITYFT-LANKVAGVQGGFLMWANDVVEDFTTNILRKDLTKID 624

Qy 630 ITIIPYIGPALNIGMILYKDFVGLIFSGAVILLEPIPEITAPVLGTALVSIVANKV 689

Db 625 VSAIPIYIGPALNISNVRGNFTFAFVTVGVTILLAPFETIPALGAFVIYSKVQERN 684

Qy 690 LTVQITDNALSRRNEKWDVEYKVIYVNNLAKVNTQIDLRKKKKEALENQAEATKAIINY 749

Db 685 EIKITDNCLEORIKRWKDSYEMMGWTLRLITQFNISIQMYDSLNYQAGAKAKIDL 744

Qy 750 QYNQYTEEEKNNINFDLSSKLANESINKAMININKFNQCSVSLNMSMTIPYGVKRIE 809

Db 745 EYKYSKSGDENIKSQVENLKNLSDVKISEAMNINKFIRECSVTYVLFKNMLPKVIDEIN 804

Qy 810 DFDASLKALLKYYIDNRCGLIGQVDRKDKVNNLTSTDIPOLSKYVDNQRLSTFTFY 869

Db 805 EFDRTKAKLILNDSHNLIIVGEVDRKAKVNSQNTIPFNISFYTNNSLKKIINEY 864

Qy 870 IKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESS----KI 924

Db 865 FNNDSKILSLQNRKNTLVDISYNAEVEEGDQLNPFPD--FKLGSSGEDRGV 921

Qy 925 EVILKNAIVNMYENFSTFWIRIPKYPNSISLNEYTIINCMMNNSGWKVSILNYGEII 984

Db 922 IVTQENIVNWSYESSFSFWIRINKWVSNLP--GYTIIDSVKNNSGWSIGIISNPLV 978

Qy 985 WILQDTQEQIKORVVKYQOMINISDYINERWIFVTITNNRLNLSKIYINGRLIDOKPISNL 1044

Db 979 FTLKQNEDEQSFNSFSDISNNAPGY-NKWFFVTNNMGNKVIYINGKLDTIKVKEL 1037

Qy 1045 GNTHASNNMFKLDGCRDT-----HRYIWIKYENFLFDEKELNEKEIKDLYDNQNSGI 1096

Db 1038 TGINFSEKTIITFEINKIPDGLITSDSDINNMWIRDYIFAKELDGKDINILFNSLQYTNV 1097

Qy 1097 LKDFWGDYIYQDKPYMLNLYDPNKTVVNVNGVIRGMYLKGPRGSVMTNLYNLSLYR 1156

Db 1098 VKDYNGDLRYNKEYYVNVNIDYLRNMYANS-----RQIVFNTR-RNNNDENE 1144

Qy 1157 GTFKFIKKYASGNKONIVRNDRVYINVVVNKEYRL-----ATNASQAQVEKILSA 1208

Db 1145 GYKIIIRKIRGNTNDRVRGDIILYFDMMINKAYNLFMKNETMYADNHSTEDIYALGR 1204

Qy 1209 LEIPDVGNLSQVVMKSKNDQGITNKC-KXNLDNNGNDIGFTGFHFQ-----NNI 1258

Db 1205 EQTKDINDNIIFOIQPMNNTYYVASQIFKSNFNGENISGTSIGTYRFRLGSDWYRHNVL 1264

Qy 1259 AKLVASWYNRQIERSRSLTGCSEWEPVDD 1289

Db 1265 VPTVKGQNVASLLESTs-----THWGFVPVSE 1291

RESULT 15

S48110

neurotoxin type F - Clostridium botulinum (fragment)

C:Species: Clostridium botulinum

C>Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C:Accession: S48110

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulin neurotoxin gene and specific ide

A:Reference number: S48103; MUID:94013372; PMID:8408542

A:Accession: S48110

A>Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 1-366 <CAM>

A:Cross-references: UNIPROT:Q57236; EMBL:X70821; NID:G407792; PIDN:CAA50152.1; PID:G40779

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 12.4%; Score 847; DB 2; Length 366;

Best Local Similarity 43.7%; Pred. No. 3.7e-30;

Matches 159; Conservative 85; Mismatches 114; Indels 6; Gaps 3;

Qy 649 KDPFVGLIFSGAVILLEPIPEITAPVLGTALVSITYA---NKVLTQITDNALSKEKNEK 705

Db 3 KENFKEAFELGAGILLFVPELLIPTILVFTKISFGSSSENKNKIKAINNLSMERETK 62

Qy	706	WDEVYKIIVTWIAKVNTOIDLIRKWKKEALENOAEATKAIINYQNOYTEEEKNNI--N	763
Db	63	WKEIYSWIVSNWLTTRINTQFNKRKEQWYQALQNVDAIKTVIEYKYNNTSDERNRLESE	122
Qy	764	FNIDDLSSKLNESINKAMININKFLNOCVSYLMNSMIPYGVKRLDPDASLKDALLKYI	823
Db	123	YNINNIREELNKKVSLAMENIERPITESSIFYLMKLINEAKVSKJREYDEGVKEYLLDYI	182
Qy	824	YDNRGTLIGQVDRLKDKVNNVTLSTDIPFOLSKYVDNQORLLSTFEYIKNIINTSIILNRY	883
Db	183	SEHSILGNSVQELNDLVSTLANSIPFELSSVTNDKILILYFNKLYKKIKDNSILDMRY	242
Qy	884	ESNHLIDLSRYASKINIGSKVNFDPIDKNOIQLFNLESSKIEVILKNAIVVNSMYENFST	943
Db	243	ENNKFDISGYGNSINGDVYIYSTVRNQGIYSSKPSSEVNIAQNNDIIYNGRYQNFSI	302
Qy	944	SPWIRIPKYENSISLANEYTIINCM-ENNSGKVSILNYGEIITWLODQOEIKORVVEKYS	1002
Db	303	SPWVRIPKYFNKVLNNEYTIIDCIRNNSGKISLNYNKIITWLODTAGNNOKLVFNNT	362
Qy	1003	QMIN 1006	
Db	363	QMIS 366	

Search completed: January 31, 2005, 14:09:14
Job time : 65 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 14:01:03 ; Search time 224 Seconds
(without alignments)
3344.365 Million cell updates/sec

Title: US-10-715-810-4
Perfect score: 6821
Sequence: 1 MPFVNQFNYPKDPVNGVDIA.....EFIPVDDGWERPLHHHHH 1302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6768	99.2	1296	2	AA06331 clostridi
2	6768	99.2	1296	2	AAQ06331 clostridi
3	6759	99.1	1295	1	BXA1_CLOBO
4	6153	90.2	1295	1	BXA2_CLOBO
5	2504.5	36.7	1268	2	Q45851 clostridium
6	2431	35.6	1278	2	Q57236 clostridium
7	2368.5	34.7	1251	2	Q9K395 clostridium
8	2339	34.3	1252	2	BAB86845 clostridi
9	2332	34.2	1252	2	Q8KZM3
10	2332	34.2	1255	2	Q9PAR6
11	2319.5	34.0	1250	1	BXE_CLOBO
12	2310	33.9	448	2	Q84G74
13	2309.5	33.9	1250	1	BXE_CLOBO
14	2296.5	33.7	1274	1	BXF_CLOBO
15	2290	33.6	1291	2	Q8GR96
16	2283.5	33.5	1280	2	Q9ZAJ5
17	2279	33.4	1296	1	BXG_CLOBO
18	2267	33.2	1291	2	Q80777
19	2266	33.2	1290	1	BXB_CLOBO
20	2258	33.1	1291	2	Q9ZAJ8
21	2258	33.1	1291	2	Q93G71
22	2255	33.1	1291	2	Q933K0
23	1890.5	27.7	1275	1	BXD_CLOBO
24	1888.5	27.7	1275	2	Q9QTG7
25	1887.5	27.7	1314	1	TETX_CLOTE
26	1854.5	27.2	1310	2	Q93N27
27	1851	27.1	1285	2	Q45967
28	1846	27.1	1285	2	Q9LBR1
29	1781.5	26.1	1280	2	Q9LBS7
30	1777.5	26.1	1280	2	Q45849
31	1774.5	26.0	1280	2	Q84163

32	1771	26.0	1291	2	Q93HT3	Q93ht3 clostridium
33	1764	25.9	1290	1	BXC1_CLOBO	P18640 clostridium
34	1360	19.9	260	2	Q45892	Q45892 clostridium
35	921.5	13.5	361	2	Q45846	Q45846 clostridium
36	918.5	13.5	361	2	Q45848	Q45848 clostridium
37	847	12.4	366	2	Q79AH9	Q79ah9 clostridium
38	818.5	12.0	441	2	Q9X708	Q9x708 clostridium
39	808.5	11.9	367	2	Q45862	Q45862 clostridium
40	805.5	11.8	367	2	Q45861	Q45861 clostridium
41	651	9.5	1198	2	O06018	O06018 clostridium
42	636	9.3	1197	2	P71117	P71117 clostridium
43	634	9.3	1197	2	O33871	O33871 clostridium
44	633.5	9.3	451	2	Q9LA13	Q9la13 clostridium
45	628	9.2	1197	2	Q9ZAJ9	Q9zaj9 clostridium

ALIGNMENTS

RESULT 1
AA06331
ID AA06331 PRELIMINARY; PRT; 1296 AA.
AC AA06331;
DT 02-MAR-2004 (Tremblrel. 27, Created)
DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)
DE BONT/A.
GN BONT/A.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hall A-byder;
RX MEDLINE=22617859; PubMed=12732962;
RA Dineen S.S., Bradshaw M., Johnson E.A.;
RT "Neurotoxin Gene Clusters in Clostridium botulinum Type A Strains:
Sequence Comparison and Evolutionary Implications.";
RL Curr. Microbiol. 46:345-352(2003).
DR EMBL: AF461540; AA06331.1;
SQ SEQUENCE 1296 AA; 149425 MW; DEAF2754AB43E6 CRC64;

Query Match	99.2%	Score	6768	DB	2	Length	1296
Best Local Similarity	99.9%	Pred.	No. 2.2e-293				
Matches	1295	Conservative	1	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MPFVNQFNYPKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIERDTFTNPEEGDLN	60				
Db	1	MPFVNQFNYPKDPVNGVDIAIYIKIPNAGOMQPVKAFKIHKNKIWIPIERDTFTNPEEGDLN	60				
Qy	61	PPPEAKQVPVSYDSTYLSVDNEKDYKLVKGVTKLFEIYSTDIGRMLLTISIVRGIPEWGG	120				
Db	61	PPPEAKQVPVSYDSTYLSVDNEKDYKLVKGVTKLFEIYSTDIGRMLLTISIVRGIPEWGG	120				
Qy	121	STIDTELKVIDTNCINVIQPDGYSRSEELNLVIIGPSADIIQFECKSFGEHVLNLTNGY	180				
Db	121	STIDTELKVIDTNCINVIQPDGYSRSEELNLVIIGPSADIIQFECKSFGEHVLNLTNGY	180				
Qy	181	GSTQYIRFSDPTFGFEESLEVDTNPLLGAGKATDPAVTLAHLIYAGHRLYGIATNP	240				
Db	181	GSTQYIRFSDPTFGFEESLEVDTNPLLGAGKATDPAVTLAHLIYAGHRLYGIATNP	240				
Qy	241	RVFKVNTNAYEMSGLEVSPFEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA	300				
Db	241	RVFKVNTNAYEMSGLEVSPFEELRTFGCHDAKFTDSIQENEFRLYYNKKFDIASTLNKA	300				
Qy	301	KSIVGTTASIQYMKNVKPKYLLSEDTSGKFSVDKLFKDKLXKMLTEIYEDNFVKFKV	360				
Db	301	KSIVGTTASIQYMKNVKPKYLLSEDTSGKFSVDKLFKDKLXKMLTEIYEDNFVKFKV	360				
Qy	361	LNKRTYLNFDKAVPKINIVPKVNYTYDGNLNTNLAAFPNGQNTENNWFKLKQFT	420				
Db	361	LNKRTYLNFDKAVPKINIVPKVNYTYDGNLNTNLAAFPNGQNTENNWFKLKQFT	420				

Db 361 LNRKTYLNFDAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDFTNDLNKEE 480
Db 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDFTNDLNKEE 480
Qy 481 ITSDTNIEAABENISLDLIQQYIITFNPDNPNENISLSSDIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIEAABENISLDLIQQYIITFNPDNPNENISLSSDIIGOLELMPNIEFPNG 540
Qy 541 KYELDKYTMHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 541 KYELDKYTMHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPNALNIGMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPNALNIGMLYKDDFVGCALIFSG 660
Qy 661 AVILLEFIPEIAPVLGTFALVSYIANKVLTVOITDNLASKRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEFIPEIAPVLGTFALVSYIANKVLTVOITDNLASKRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTGIGVDRLKDK 840
Db 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTGIGVDRLKDK 840
Qy 841 VNNTLSTDIPQLSKYVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Db 841 VNNTLSTDIPQLSKYVDNORLLSTFTTEYIKNIINTSILNLRYESNHLIDLRYASKINI 900
Qy 901 GSKVNFDPIDKXQIQLFNLESKIEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Db 901 GSKVNFDPIDKXQIQLFNLESKIEVILKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
Qy 961 EYTIINCENNSGKWSLNYGEIITWLTQDTEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
Db 961 EYTIINCENNSGKWSLNYGEIITWLTQDTEIKQKRVVFKYSQMINISDYINRWIFVTIT 1020
Qy 1021 NNRNNSKIYNGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Db 1021 NNRNNSKIYNGRLIDQKPIISNLGNHASNNIMFKLDGCRDTHRYIWKYFNLFDKELN 1080
Qy 1081 EKEIKDLYDNOSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNVNGIRGYMLKGP 1140
Db 1081 EKEIKDLYDNOSGILKDFWGDYLOYDKPYMLNLYDPNKYVDVNVNGIRGYMLKGP 1140
Qy 1141 GSVMTNIIYLSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Db 1141 GSVMTNIIYLSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
Qy 1201 GVEKILSALETDPVGNLSQVVMKSKNDQGTNCKMNLQDNNGNDIGFTGHGFNNIAK 1260
Db 1201 GVEKILSALETDPVGNLSQVVMKSKNDQGTNCKMNLQDNNGNDIGFTGHGFNNIAK 1260
Qy 1261 LVASNNYNRQIERSRRTIGCSWEFIPVDDGGERPL 1296
Db 1261 LVASNNYNRQIERSRRTIGCSWEFIPVDDGGERPL 1296

RESULT 2

AAQ06331 PRELIMINARY; PRT; 1296 AA.
AC AAQ06331;
DT 02-MAR-2004 (TremBLrel. 27, Created)
DT 02-MAR-2004 (TremBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TremBLrel. 27, Last annotation update)
DE Neurotoxin BoNT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Allergan-Hall A;
RX MEDLINE=22919384; PubMed=14557061;
RA Zhang L., Lin W.J., Li S., Aoki K.R.;
RT "Complete DNA sequences of the botulinum neurotoxin complex of
Clostridium botulinum type A-Hall (Allergan) strain.";
RL Gene 315:21-32(2003).
DR EMBL; AF488749; AAQ06331.1; -.
KW Neurotoxin.
SQ SEQUENCE 1296 AA; 149425 MW; DEACF2754AE43E6 CRC64;

Query Match 99.2%; Score 6768; DB 2; Length 1296;

Best Local Similarity 99.9%; Pred. No. 2.2e-293;

Matches 1295; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPFVNKQFNKYKDPVNGVDIAYIKIPNAGQMPKAFKIHNKIWIIPERDFTTNPEDGLN 60
Db 1 MPFVNKQFNKYKDPVNGVDIAYIKIPNAGQMPKAFKIHNKIWIIPERDFTTNPEDGLN 60
Qy 61 PPPEAKQVPVSYVDSTVLTSDNEKDNVYKGVTKLPERIYSTDLGRMLLTISVIRGIPWGG 120
Db 61 PPPEAKQVPVSYVDSTVLTSDNEKDNVYKGVTKLPERIYSTDLGRMLLTISVIRGIPWGG 120
Qy 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
Db 121 STIDTELKVIDTNCINVIQPDGYSRSEELNVIIGPSADIIQPECKSFGEHVLNLTNGY 180
Qy 181 GSTQYIRFSPDFTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Db 181 GSTQYIRFSPDFTFGPEESLEVDTNPLLGAGKATDPAVTLAHELIYAGHRLYGIAPNP 240
Qy 241 RVKQVNTNAYENSGLEVSEELRTFGCHDAKFTDSLOENEFRLYYNKKFDLASTLKA 300
Db 241 RVKQVNTNAYENSGLEVSEELRTFGCHDAKFTDSLOENEFRLYYNKKFDLASTLKA 300
Qy 301 KSVIGTTASIQYMKNVFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYEDNFVKFKV 360
Db 301 KSVIGTTASIQYMKNVFKEKYLLEDSTSGKFSVDKLFKDKLYKMLTEIYEDNFVKFKV 360
Qy 361 LNRKTYLNFDAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Db 361 LNRKTYLNFDAVFKINIVPKVNTIYDGFNLRNTNLAANFNGQNTENNMFVKLNFT 420
Qy 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDFTNDLNKEE 480
Db 421 GLFEFYKLLCVRGIIITSKSLDGYNKALNDLCIKVNNWDLFFSPSEDFTNDLNKEE 480
Qy 481 ITSDTNIEAABENISLDLIQQYIITFNPDNPNENISLSSDIIGOLELMPNIEFPNG 540
Db 481 ITSDTNIEAABENISLDLIQQYIITFNPDNPNENISLSSDIIGOLELMPNIEFPNG 540
Qy 541 KYELDKYTMHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Db 541 KYELDKYTMHYLRAQEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKKYKATEA 600
Qy 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPNALNIGMLYKDDFVGCALIFSG 660
Db 601 AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIPIYGPNALNIGMLYKDDFVGCALIFSG 660
Qy 661 AVILLEFIPEIAPVLGTFALVSYIANKVLTVOITDNLASKRNEKWDDEVYKIVTNWLAK 720
Db 661 AVILLEFIPEIAPVLGTFALVSYIANKVLTVOITDNLASKRNEKWDDEVYKIVTNWLAK 720
Qy 721 VNTQIDILIRKMKKEALENQAATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
Db 721 VNTQIDILIRKMKKEALENQAATKAIINYQNYQTEEEKNNINFNIDDLSSKLNESINKA 780
Qy 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTGIGVDRLKDK 840
Db 781 MININKFLNOCVSYLMNSMIPYGVKRLDFDASLKDALLKYIYDNRGTGIGVDRLKDK 840

Qy 841 VVNTLSTDDIPOLSKYVDNQRLSTFTTEYIKNIINTSILMLRYBSNHLIDLSRYASKINI 900
 Db 841 VVNTLSTDDIPOLSKYVDNQRLSTFTTEYIKNIINTSILMLRYBSNHLIDLSRYASKINI 900
 Qy 901 GSKVNFDPIDKNQIQLFNLESSKLEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
 Db 901 GSKVNFDPIDKNQIQLFNLESSKLEVLKNAIVNSMYENFSTFWIRIPKYFNSISLNN 960
 Qy 961 EYTIINCMMNSGKWKSLNTGEEIIITWLTQDTEIKQIRVVFYKYSQMINISDYINRWIFVTIT 1020
 Db 961 EYTIINCMMNSGKWKSLNTGEEIIITWLTQDTEIKQIRVVFYKYSQMINISDYINRWIFVTIT 1020
 Qy 1021 NNRLLNSKIYINGRLIDQKPSINLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
 Db 1021 NNRLLNSKIYINGRLIDQKPSINLGNTHASNNIMFKLDGCRDTHRYIWIYFNLFDKELN 1080
 Qy 1081 EKETIKDYDQNSGIIKDFWGDYLOVDKPYVLMNLVDPNKYVDVNVNGVIRGYMYLKGPR 1140
 Db 1081 EKETIKDYDQNSGIIKDFWGDYLOVDKPYVLMNLVDPNKYVDVNVNGVIRGYMYLKGPR 1140
 Qy 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
 Db 1141 GSVMTTNIYLNSSLYRGTKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRLATNASQA 1200
 Qy 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKVNLQDNNGNDIGFTGFHFQFNNAK 1260
 Db 1201 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNCKVNLQDNNGNDIGFTGFHFQFNNAK 1260
 Qy 1261 LVASNNYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296
 Db 1261 LVASNNYNRQIERSRSLTGCSEWEPVDDGNGERPL 1296

RESULT 3
 EXAL_CLOBO STANDARD; PRT; 1295 AA.
 AC P10845; P01561; P18639;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].
 GN Name: botA; Synonyms: bna, atx;
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1491;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Type A / NCTC 2916;
 RX MEDLINE=90235864; PubMed=2185020;
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J., Shone C.C.,
 RA Atkinson T., Melling J., Minton N.P.;
 RT "The complete amino acid sequence of the Clostridium botulinum type A
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
 RT gene.";
 RL Eur. J. Biochem. 189:73-81(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Type A / 62A;
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz B., Kuarzono H., Willie M., Frevent J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 RT with other clostridial neurotoxins.";
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [3]
 RN SEQUENCE OF 1-65 FROM N.A.
 RC STRAIN=Type A / 62A;
 RX MEDLINE=97016817; PubMed=8863443;
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
 RT "Organization and phylogenetic interrelationships of genes encoding

RT components of the botulinum toxin complex in proteolytic Clostridium
 RT botulinum types A, B, and F: evidence of chimeric sequences in the
 RT gene encoding the nontoxic nonhemagglutinin component.";
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
 RN [4]
 RN SEQUENCE OF 1-34 FROM N.A.
 RC STRAIN=Type A / Hall;
 RX MEDLINE=89350959; PubMed=2669749;
 RA Betley M.J., Somers E., Dasgupta B.R.;
 RT "Characterization of botulinum type A neurotoxin gene: delineation of
 RT the N-terminal encoding region.";
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
 RN [5]
 RN SEQUENCE OF 1-18 FROM N.A.
 RC STRAIN=Type A / NIH;
 RX MEDLINE=96096783; PubMed=8521962;
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
 RT components of Clostridium botulinum type A progenitor toxins.";
 RL FEBS Lett. 376:41-44(1995).
 RN [6]
 RN SEQUENCE OF 1-16.
 RX MEDLINE=84178501; PubMed=6370252;
 RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
 RT "Partial amino acid sequence of the heavy and light chains of
 RT botulinum neurotoxin type A.";
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
 RN [7]
 RN SEQUENCE OF 1-46.
 RA Dasgupta B.R., Foley J., Niece R.;
 RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
 RL Biochemistry 26:4162-4162(1987).
 RN [8]
 RN SEQUENCE OF 1-5 AND 444-456.
 RX MEDLINE=91120847; PubMed=2126206;
 RA Dasgupta B.R., Dekleva M.L.;
 RT "Botulinum neurotoxin type A: sequence of amino acids at the N-
 RT terminus and around the nicking site.";
 RL Biochimie 72:661-664(1990).
 RN [9]
 RN SEQUENCE OF 448-464 AND 872-895.
 RX MEDLINE=89024662; PubMed=3178218;
 RA Sartymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
 RT halves and partial sequences.";
 RL Arch. Biochem. Biophys. 266:142-151(1988).
 RN [10]
 RN SEQUENCE OF 448-482.
 RX MEDLINE=85285016; PubMed=3896784;
 RA Shone C.C., Hambleton P., Melling J.;
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
 RT and purification of two tryptic fragments. Proteolytic action near the
 RT COOH-terminus of the heavy subunit destroys toxin-binding activity.";
 RL Eur. J. Biochem. 151:75-82(1985).
 RN [11]
 RN SEQUENCE OF 866-879 AND 1147-1218.
 RX Gimenez J.A., Dasgupta B.R.;
 RA "Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72,
 RT 45, 42, and 18 kD fragments.";
 RL J. Protein Chem. 12:351-363(1993).
 RN [12]
 RN IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94063091; PubMed=8243676;
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
 RA Benfenati F., Wilson M.C., Montecucco C.;
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
 RT COOH-terminal peptide bonds.";
 RL FEBS Lett. 335:99-103(1993).
 RN [13]
 RN IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94124495; PubMed=8294407;
 RA Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,

RA Jahn R., Niemann H.;
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins."; [14]
 RN J. Biol. Chem. 269:1617-1620(1994).
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
 RX MEDLINE=2155894; PubMed=11700044; DOI=10.1006/bbr.2001.5911;
 RA Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
 RT "Site-directed mutagenesis identifies active-site residues of the light chain of botulinum neurotoxin type a."; [15]
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
 RN X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
 RP MEDLINE=94455071; PubMed=9783750;
 RX Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
 RA "Crystal structure of botulinum neurotoxin type A and implications for toxicity."; [16]
 RL Nat. Struct. Biol. 5:898-902(1998).
 CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure.
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexcitatory apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit.
 CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for the treatment of strabismus and blepharospasm associated with dystonia and cervical dystonia. Also used for the treatment of hemifacial spasm and a number of other neurological disorders characterized by abnormal muscle contraction.
 CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
 CC -1- SIMILARITY: Belongs to peptidase family M27.
 CC -1- DATABASE: NAME=BOTOX product information Web site; WWW="http://www.botox.com/site/".
 CC -1- DATABASE: NAME=Protein Spotlight; NOTE=Issue 19 of February 2002; WWW="http://www.expasy.org/spotlight/articles/spl019.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52066; CAA36289.1; -
 CC EMBL; M30196; AAA23262.1; -
 CC EMBL; X92973; CAA63551.1; -
 CC EMBL; D67030; BAA11051.1; -
 CC EMBL; M27892; AAA23269.1; -
 CC PIR; A35294; BTCLAB.
 CC PDB; 3BTA; X-ray; A=1-1295.
 CC MEROPS; M27.002; -
 CC InterPro; IPR008985; ConA like lec_gl.
 CC InterPro; IPR011065; Kunitz like.
 CC InterPro; IPR000395; Peptidase M27.
 CC InterPro; IPR006025; Pept M Zn BS.
 CC Pfam; PF01742; Peptidase M27; 1.
 CC PRINTS; PR00760; BONTOXILYSIN.
 CC ProDom; PD001963; Bontoxilysin; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC 3D-structure; Direct protein sequencing; Hydrolase; Metalloprotease;

KW Neurotoxin; Pharmaceutical; Transmembrane; Zinc.
 FT INIT MET 0
 FT CHAIN 1 447 Botulinum neurotoxin A light-chain.
 FT CHAIN 448 1295 Botulinum neurotoxin A heavy-chain.
 FT METAL 222 222 Zinc (catalytic).
 FT ACT_SITE 223 223 Zinc (catalytic).
 FT METAL 226 261 Zinc (catalytic).
 FT METAL 261 261 Zinc (catalytic).
 FT DISULFID 429 453 Interchain.
 FT TRANSMEM 1234 1279 Potential.
 FT TRANSMEM 626 646 Potential.
 FT VARIANT 26 26 V -> A.
 FT MUTAGEN 261 261 E->A: Drastic decrease in enzymatic activity.
 FT MUTAGEN 265 265 F->A: Decreases enzymatic activity.
 FT MUTAGEN 365 365 Y->A: Decreases enzymatic activity.
 Query Match 99.1%; Score 6759; DB 1; Length 1295;
 Best Local Similarity 99.8%; Pred. No. 5.5e-293;
 Matches 1293; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PFVKQFNKYKDPVNGVDIAVIKIPNAGQMPQVAFKIHKKIWIPIPRDPTNPEEGDLNP 61
 DB 1 PFVKQFNKYKDPVNGVDIAVIKIPNAGQMPQVAFKIHKKIWIPIPRDPTNPEEGDLNP 60
 QY 62 PPEAKQVPVSYDYSTYLSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTIVRGIPFWGGS 121
 DB 61 PPEAKQVPVSYDYSTYLSTDNKDNLYKGVTKLPERIYSTDLGRMLLTSTIVRGIPFWGGS 120
 QY 122 TIDTELKVIDTNCINVIQDPGYSRSEELNLIIVIGPSADIIQFCKSFHGVMLNTRNGY 181
 DB 121 TIDTELKVIDTNCINVIQDPGYSRSEELNLIIVIGPSADIIQFCKSFHGVMLNTRNGY 180
 QY 182 STQVIRFSPDFTGRFEESLEVDNPLLGAGKFAIDPAVTLAHLIYAGHLYGIAINPNR 241
 DB 181 STQVIRFSPDFTGRFEESLEVDNPLLGAGKFAIDPAVTLAHLIYAGHLYGIAINPNR 240
 QY 242 VFQVNTNAYEMSGLEVSPFEELRTFGHDAKFDLSQENEFRLYYNKKFKDIASTLNKAK 301
 DB 241 VFQVNTNAYEMSGLEVSPFEELRTFGHDAKFDLSQENEFRLYYNKKFKDIASTLNKAK 300
 QY 302 SIQVTTASLOQYMKVPEKYLKLLSEDTSGKFSVDKLPDKLYKMLTEIYEDNFVKPFVL 361
 DB 301 SIQVTTASLOQYMKVPEKYLKLLSEDTSGKFSVDKLPDKLYKMLTEIYEDNFVKPFVL 360
 QY 362 NRKTYLNFDKAVKINIVPKVNTIYDGFNLNLTNLAANFNGQNTNINNFTKLNFTG 421
 DB 361 NRKTYLNFDKAVKINIVPKVNTIYDGFNLNLTNLAANFNGQNTNINNFTKLNFTG 420
 QY 422 LPFQYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEEI 481
 DB 421 LPFQYKLLCVRGITTSKTSKSLDKGYNKALNDLCIKVNNWDLFPSPSEDNFTDLNKGEEI 480
 QY 482 TSDNTNTEAAEENISLDLIQOYLYTFNFDNPEINISTENISLSSDIIGQLELMPNTERFENGK 541
 DB 481 TSDNTNTEAAEENISLDLIQOYLYTFNFDNPEINISTENISLSSDIIGQLELMPNTERFENGK 540
 QY 542 KYELDQYTMFHYLRAOEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 601
 DB 541 KYELDQYTMFHYLRAOEFHGKSRIALTNSVNEALLNPSRVYTFSSDYVKVKNKATEAA 600
 QY 602 MFLGWVQLVYDFTDETSFVSTTKDIADITIIPIYIGPALNIGMLYKDDFVGALIFSGA 661
 DB 601 MFLGWVQLVYDFTDETSFVSTTKDIADITIIPIYIGPALNIGMLYKDDFVGALIFSGA 660
 QY 662 VILLEFPEITAIPIVLTGTFALSVIANKVLVQTDNALSQRNEKWDVYKIVTNWLAKV 721
 DB 661 VILLEFPEITAIPIVLTGTFALSVIANKVLVQTDNALSQRNEKWDVYKIVTNWLAKV 720
 QY 722 NTQIDLIRKQKALENOAEATKAIINYQVNTYEEKNNINFNIDDLSSKLINESINKAM 781
 DB 721 NTQIDLIRKQKALENOAEATKAIINYQVNTYEEKNNINFNIDDLSSKLINESINKAM 780

QY 782 ININKELNQC SVSYLMSMIPYGVRLLEDPSALKDALLKYIYVNRGTLIGQVDRDKKV 841
 DB 781 ININKELNQC SVSYLMSMIPYGVRLLEDPSALKDALLKYIYVNRGTLIGQVDRDKKV 840
 QY 842 NNTLSTDPQLSKYVDNQLSTFEYIKNIINTSILNRYESNHLIDLRYASKINIG 901
 DB 841 NNTLSTDPQLSKYVDNQLSTFEYIKNIINTSILNRYESNHLIDLRYASKINIG 900
 QY 902 SKVNFDPIDKQQLFNLESSKIEVLKNAIVNYSNMFSTFWIRPKYFNSISLNE 961
 DB 901 SKVNFDPIDKQQLFNLESSKIEVLKNAIVNYSNMFSTFWIRPKYFNSISLNE 960
 QY 962 YTIINCWENNSGKWSLNYGEIITWLODQEIQRVVFVKYSOMINSDYINRWIFVTITN 1021
 DB 961 YTIINCWENNSGKWSLNYGEIITWLODQEIQRVVFVKYSOMINSDYINRWIFVTITN 1020
 QY 1022 NRLNNSKIYNGRLIDOKPIISNIGNTHASNIMFKLDCGDRTHRYIWKYFNLFDKELNE 1081
 DB 1021 NRLNNSKIYNGRLIDOKPIISNIGNTHASNIMFKLDCGDRTHRYIWKYFNLFDKELNE 1080
 QY 1082 KEIKDYDNGNSGILKDFWGDYLDYDKPYMLNLYDPNKYVDVNVNNGIRGYMLKGPGRG 1141
 DB 1081 KEIKDYDNGNSGILKDFWGDYLDYDKPYMLNLYDPNKYVDVNVNNGIRGYMLKGPGRG 1140
 QY 1142 SVMTTNIYLSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKYKRYLATNASQAG 1201
 DB 1141 SVMTTNIYLSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKYKRYLATNASQAG 1200
 QY 1202 VEKILSALETPDVGNSLQVVMKSKNDQGTNCKMKNLQDNGDNGDIFGIFHGFNNIAKL 1261
 DB 1201 VEKILSALETPDVGNSLQVVMKSKNDQGTNCKMKNLQDNGDNGDIFGIFHGFNNIAKL 1260
 QY 1262 VASWYNRQIERSRSLGCSWEFIPVDGNGERPL 1296
 DB 1261 VASWYNRQIERSRSLGCSWEFIPVDGNGERPL 1295

RESULT 4
 BXA2_CLOBO STANDARD; PRT; 1295 AA.
 AC Q45894; P77780;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].
 DE Name=botA; Synonyms=bna, atx;
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Type A / Kyoto-F;
 RX MEDLINE=94143603; PubMed=8310180;
 RA Wallens A., East A.K., Lawson P.A., Collins M.D.;
 RT "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins.";
 RL Res. Microbiol. 144:547-556(1993).
 RN [2]
 RP SEQUENCE OF 1-65 FROM N.A.
 RC STRAIN=Type A / Kyoto-F;
 RX MEDLINE=97016817; PubMed=8863443;
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
 RT "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the RT gene encoding the nontoxic nonhemagglutinin component.";
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
 CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin

CC binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure (By similarity).
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.
 CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H) (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
 CC -1- SIMILARITY: Belongs to peptidase family M27.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X73423; CAA51824.1; -;
 CC ENBL: X87974; CAA61234.1; -;
 CC PIR: I40645; I40645.
 CC PDB: 1E1H; X-ray; A/C=1-249, B/D=250-423.
 CC MEROPS: M27.002; -;
 CC InterPro: IPR008985; ConA like lec_g1.
 CC InterPro: IPR011065; Kunitz like.
 CC InterPro: IPR000395; Peptidase M27.
 CC InterPro: IPR006025; Pept_M_Zn_BS.
 CC Pfam: PF01742; Peptidase M27; 1.
 CC PRINTS: PR00760; BONTOXILYSIN.
 CC ProDom: PD001963; Bontoxilysin; 1.
 CC PROSITE: PS00142; ZINC PROTEASE; FALSE NEG.
 CC 3D-structure: Hydrolase; Metalloprotease; Neurotoxin; Transmembrane; Zinc.
 CC INIT_MET 0 0 By similarity.
 CC CHAIN 1 447 Botulinum neurotoxin A light-chain.
 CC CHAIN 448 1295 Botulinum neurotoxin A heavy-chain.
 CC METAL 222 222 Zinc (catalytic) (By similarity).
 CC ACT_SITE 223 223 Zinc (catalytic) (By similarity).
 CC METAL 226 226 Zinc (catalytic) (By similarity).
 CC DISULFID 429 453 Interchain (By similarity).
 CC TRANSMEM 1234 1279 By similarity.
 CC TRANSMEM 626 646 Potential.
 CC SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;
 QY Query Match 90.2%; Score 6153; DB 1; Length 1295;
 DB Best Local Similarity 89.9%; Pred. No. 5.7e-266;
 QY Matches 1164; Conservative 66; Mismatches 65; Indels 0; Gaps 0;
 QY 2 PFVKNQFNYKDPVNGVDIAYIKIPNAGOMQPVKAFKHKNKIWIIPERDTFTNPEEGDLP 61
 DB 1 PFVKNQFNYKDPVNGVDIAYIKIPNAGOMQPVKAFKHKNKIWIIPERDTFTNPEEGDLP 60
 QY 62 PPEAKQVPVSYDYSTVLTSTNEKDNVLYKGVTKLPERIYSTDLGRMLTSTVIRGPFWGG 121
 DB 61 PPEAKQVPVSYDYSTVLTSTNEKDNVLYKGVTKLPERIYSTDLGRMLTSTVIRGPFWGG 120
 QY 122 TIDTELKVIDTNCINVIQDPGYSRSEELNVLIIQPSADIIIOFECKSFGEVNLNTRNGY 181
 DB 121 TIDTELKVIDTNCINVIQDPGYSRSEELNVLIIQPSADIIIOFECKSFGEVNLNTRNGY 180
 QY 182 STQVIRSPDFTGPFRESLEVDNPNLLGAGKPAITDPAVTLAHELIYAGHLYGTAIPNR 241
 DB 181 STQVIRSPDFTGPFRESLEVDNPNLLGAGKPAITDPAVTLAHELIYAGHLYGTAIPNR 241

Db 181 STQVIRSPDFTFGFBESLEVDNTPLLGAGKFAATDPVTLAHELIHAERLYGAINPNR 240
Qy 242 VFKNNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFKDIASTLNKAK 301
Db 241 VFKNNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKKFKDIASTLNKAK 300
Qy 302 SIQVTTASLQVMKNVFKKYLSSDTSKFSVDKLFKDKLYKMLTEIYTDNFVFKFKVL 361
Db 301 SIQVTTASLQVMKNVFKKYLSSDTSKFSVDKLFKDKLYKMLTEIYTDNFVFKFKVI 360
Qy 362 NRKTYLNFDAKAVKINIVPKVNTIYDGFNLRTNLAANFNGQNTNINMNFYKLNKFTG 421
Db 361 NRKTYLNFDAKAVKINIVPKVNTIYDGFNLRTNLAANFNGQNTNINMNFYKLNKFTG 420
Qy 422 LFEFYKLLCVRGITTSKTSKLDKGNKALNDLCIKVNNWDLFPSPSDNFTNDLNGKEEI 481
Db 421 LFEFYKLLCVRGITTSKTSKLDKGNKALNDLCIKVNNWDLFPSPSDNFTNDLNGKEEI 480
Qy 482 TSDTNIEAAENISLDDLIQOYLLTFNDFNBPENISIEENLSSDIIGOLELMPNIEFPNGK 541
Db 481 TADTNIEAAENISLDDLIQOYLLTFNDFNBPENISIEENLSSDIIGOLELMPNIEFPNGK 540
Qy 542 KYELDKYTMFHYLRAQFEHGKSHALNTSNVNEALLNPSRVYTFSSDYKKNKATEAA 601
Db 541 KYELDKYTMFHYLRAQFEHGKSHALNTSNVNEALLNPSRVYTFSSDYKKNKATEAA 600
Qy 602 MFLGWVQLVYDFTDSEVSTTDKIADITIIPIYIGPALNIGMLKDDFVGALIFSGA 661
Db 601 MFLGWVQLVYDFTDSEVSTTDKIADITIIPIYIGPALNIGMLKDDFVGALIFSGA 660
Qy 662 VILLEFIPEALPVGTFALVSYANKVLVTQTTIDNALSKEKDEWYKYIYVNNLAKV 721
Db 661 VILLEFIPEALPVGTFALVSYANKVLVTQTTIDNALSKEKDEWYKYIYVNNLAKV 720
Qy 722 NTQIDLRKMKALENOAETAKIINYQNYTEEEKNNINFNIDLSKLNESINKAM 781
Db 721 NTQIDLRKMKALENOAETAKIINYQNYTEEEKNNINFNIDLSKLNESINKAM 780
Qy 782 ININKFNLQCSVSLNMSLPIYGVKRLDPEASUKDALLKYIYDNRGTGLQVDRKDKV 841
Db 781 ININKFNLQCSVSLNMSLPIYGVKRLDPEASUKDALLKYIYDNRGTGLQVDRKDKV 840
Qy 842 NNTLSTDIPQLSKYVDNRLSTFTYIKNIINTSILNIRYESNHLIDLSRYASKINIG 901
Db 841 NNTLSTDIPQLSKYVDNRLSTFTYIKNIINTSILNIRYESNHLIDLSRYASKINIG 900
Qy 902 SKYNFDPIDKNOIOLFNLESSKTEVILKNAIVNMSYENFSTSWIPIKYFNSISLANE 961
Db 901 SKYNFDPIDKNOIOLFNLESSKTEVILKNAIVNMSYENFSTSWIPIKYFNSISLANE 960
Qy 962 YTIINCENNSGKVSILNYGIEIWTLODQTOBKORVVFVFKYSQMINISDIYNRMIFVTITN 1021
Db 961 YTIINCENNSGKVSILNYGIEIWTLODQTOBKORVVFVFKYSQMINISDIYNRMIFVTITN 1020
Qy 1022 NRLNNSKIYINGRLIDOKPISNLNIIHASNNIMFKLDCGRDTHRYIWKYFNLFDKELNE 1081
Db 1021 NRLNNSKIYINGRLIDOKPISNLNIIHASNNIMFKLDCGRDTHRYIWKYFNLFDKELNE 1080
Qy 1082 KEIKOLDYDQNSGILKDFWGDYLODKPYMMLNLYDPNKYVDVNNVGIKGYMLKGPGRG 1141
Db 1081 KEIKOLDYDQNSGILKDFWGDYLODKPYMMLNLYDPNKYVDVNNVGIKGYMLKGPGRG 1140
Qy 1142 SVMNTNIIYNSSLYRGTKFIKKYASGNKONIYVNNDRVYVNNVKNKEVRLATNASQAG 1201
Db 1141 SVMNTNIIYNSSLYRGTKFIKKYASGNKONIYVNNDRVYVNNVKNKEVRLATNASQAG 1200
Qy 1202 VEKILSALEIPDVGNLSQVVMKSKNDQGTNCKMKNLDQNNGNDIGFIFGHQNNIAKL 1261
Db 1201 VEKILSALEIPDVGNLSQVVMKSKNDQGTNCKMKNLDQNNGNDIGFIFGHQNNIAKL 1260
Qy 1262 VASNNYRQIERSRRTGCSWEFIPVDDGGERPL 1296
Db 1261 VASNNYRQIERSRRTGCSWEFIPVDDGGERPL 1295

RESULT 5
Q45851 PRELIMINARY; PRT; 1268 AA.
ID Q45851
AC Q45851;
DT 01-NOV-1996 (TREMUREL. 01, Created)
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)
DT 01-MAR-2004 (TREMUREL. 26, Last annotation update)
DE Neurotoxin type F.
GN Name:bont /i;
OS Clostridium baratii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
neurotoxin: comparison with other clostridial neurotoxins";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -;
DR PIR; S33411; S33411.
DR HSSP; Q45894; IE1H.
DR MEROPS; M27.002; -;
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009405; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR011065; Kunitz like.
DR InterPro; IPR000395; Peptidase_M27.
DR InterPro; IPR006025; Pept_M_zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR PRODOM; PD001963; Botulinum; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1268 AA; 145512 MW; 963040091AC15ED2 CRC64;
Query Match 36.7%; Score 2504.5; DB 2; Length 1268;
Best Local Similarity 41.4%; Pred. No. 2.4e-103;
Matches 552; Conservative 242; Mismatches 433; Indels 107; Gaps 36;
Qy 1 MPVFNQFNKQVGVNDIAYIKIP-NAGQQPVKAFKHKHWIPIPERDTFTNPERGDL 59
Db 1 MPVFNQFNKQVGVNDIAYIKIP-NAGQQPVKAFKHKHWIPIPERDTFTNPERGDL 59
Qy 60 NPPEAKQVPSVYDSTYSLTDNEKNYKGVTKLFIERYISTDLGRMLLTSIVRGIPFWG 119
Db 60 NPPEAKQVPSVYDSTYSLTDNEKNYKGVTKLFIERYISTDLGRMLLTSIVRGIPFWG 119
Qy 120 GSTIDTELKVIDTNCIN-----VIQDGSYSSEELNVLIGPSADIIOPECKSPGHEV 172
Db 120 GSTIDTELKVIDTNCIN-----VIQDGSYSSEELNVLIGPSADIIOPECKSPGHEV 172
Qy 173 L-----NLTRNGVGSYQYIRFSPDFTFGFEESLEVDTNLLGAGKFAATPAVTLAHELI 226
Db 173 L-----NLTRNGVGSYQYIRFSPDFTFGFEESLEVDTNLLGAGKFAATPAVTLAHELI 226
Qy 227 YAGHRLYGI-AINPNRVKNTNAYVEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLY 285
Db 227 YAGHRLYGI-AINPNRVKNTNAYVEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLY 285
Qy 286 YNKKFKDIASTL---NKAKSIVGTTASLQVMKNVFKKYLSSDTSKFSVDKLFKDKLY 342
Db 286 YNKKFKDIASTL---NKAKSIVGTTASLQVMKNVFKKYLSSDTSKFSVDKLFKDKLY 342
Qy 343 KMLTEIYTDNFVKFKVNLKTYLNFDAKAVKINIVPKVNTIYDGFNLRTNLAANFN 402
Db 343 KMLTEIYTDNFVKFKVNLKTYLNFDAKAVKINIVPKVNTIYDGFNLRTNLAANFN 402
Qy 394 KKLFS-FTCEDLAQKQVKNRNYLFHFKPFRLLDLDLDNIYSISBGFNI--GSLRVNNN 394
Db 394 KKLFS-FTCEDLAQKQVKNRNYLFHFKPFRLLDLDLDNIYSISBGFNI--GSLRVNNN 394


```
Db 447 RVNRRELFVASESSYNNENDINTPKEDIDTTNNNNYRN-NLD---EVILDYSETIPQ- 501
Qy 515 ISIEINLSSDIIGOLELMPNTERFPNGKYEYLDKY-----TWFYHLYRAQEFEGHRSIALT 569
Db 502 ISNQTALT-LVQDDSYYP---RYDSNGTSELEHNVDLVNFFVFLHQAQYGEGETNISLT 557
Qy 570 NSVNEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWVQLVYDFTDTESEVSTDKIAD 629
Db 558 SSIDTALSESQVYTFSSSEFINTINKPVHAALFISWINQVIRDTFTTEATQKSTFDKIAD 617
Qy 630 ITIILPIGPAINLCNMLYKDDPFGALFSGAVILLFPIEAIAPVLGTALVSYIA--- 686
Db 618 ISLVVPVGLALNTGNEQKFEAFELLAGILLFVPELLPTTLVLTFIKSFISGE 677
Qy 687 NKVLTVOTIDNALSKRNEKDEVDVYKIVTNWLAKVNTQIDILIRKKMKEALENOAEATKAI 746
Db 678 NKNKIIKAINSLMERETKWEIYSWISVSNWLTRINTQFNKRKEQMYQALONQVDAIKTV 737
Qy 747 INYOQNYTEEEKNI--NFNIDDLSSKLINESINKAMININKFNQCSVSYLNMSTMIPYG 804
Db 738 IEYKYNVTSDERNLRSEYNNIREELNKKVSLAMENIERFTESSIFLYMLKLINEAK 797
Qy 805 VKRLEDFDASIKDALLKXIYDNRGTLIGQVDRLKXKVNLTSTDIPLQSKYVDNQRLLS 864
Db 798 VSKLREYDEGVKYLDDYISHRSLGNSVQELNDLVTSTLNNSSIPPELSSYNTDKILIL 857
Qy 865 TFEYIKNIINTSILNLRYSNHLIDLSRVASKINIGSKVNFDPIDKNQIQLFNLESSKI 924
Db 858 YFNKLYKKIKONSLIDRWYENKEDIDISGYSNISINGDVYIYSTNRNQFIYSSKPEV 917
Qy 925 EVILKNAIVNSMYENSTFWIRPKYFNSISLANEYTIINCW-ENNSGKVKSLNYGEI 983
Db 918 NIAQNNDIYNGRYQNFSPWIRPKYFNKVNLANEYTIIDCIIRNNNSGKISLNYNKI 977
Qy 984 IWTLODQOEIKORVVKYSOMINISDVNIWFIWITNNRLNSKIYINGELIDOKPISN 1043
Db 978 IWTLODQAGNKLQKLVNFYQMSISDYINKWIFWITNNRLNSKIYINGELIDOKPISN 1037
Qy 1044 LGNIHASNIMFKLDGCRDTHRYIWKYFNLFDKELNEKEIKOLYDQNSGILKDFWGD 1103
Db 1038 LGDIHVSNDILFKIVGNDT-RYVGIRYFKVDFDELGTETIETLYSDEPDPSILKDFWGN 1096
Qy 1104 YLQVDKPYMLNLYDPNKNYVDVNVGIRGMYLWYKGRPGSVWNTNINLSSLYRGTKEIK 1163
Db 1097 YLLYNKRYLLNLLRTDKSITQNS---NFLNINQORGVYQKPNIFSNTLYTGVFVIIR 1152
Qy 1164 KYAS---GNKDNIVRNDRVYINNVKNKEVRLATNASQAVEKILSALEIPDVG-N-LSQ 1219
Db 1153 KNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKPEKILKLRITSNNSNLGQ 1212
Qy 1220 VVWMSKNDQGITNKKCNLQDNGNDIGFTGFHFQFNIAKLVASNNWYNRQIERSSRTLQ 1279
Db 1213 IIVWDS-----IGNCTWTFQNNNGNGLLGFHSNN---LVASSWYNNIRKNTSSNG 1263
Qy 1280 CSMEFIPVDGQGE 1293
Db 1264 CFWSFISKEHGWOE 1277

RESULT 7
Q9K395
ID Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Type E botulinum toxin.
GN Name=Bot/E;
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1492;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 095, LCL 155, KZ 1899, KZ 1897, KZ 1898, KZ 1886, KZ 1887,
KZ 1889, KZ 1890, KZ 1891, and LCL 063;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karsawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT "Genetic analysis of type E botulinum toxin-producing Clostridium
butyricum strains.";
RL Appl. Environ. Microbiol. 66:4992-4997 (2000).
DR EMBL; AB037714; BAB03522.1; -
DR EMBL; AB037704; BAB03512.1; -
DR EMBL; AB037705; BAB03513.1; -
DR EMBL; AB037706; BAB03514.1; -
DR EMBL; AB037707; BAB03515.1; -
DR EMBL; AB037708; BAB03516.1; -
DR EMBL; AB037709; BAB03517.1; -
DR EMBL; AB037710; BAB03518.1; -
DR EMBL; AB037711; BAB03519.1; -
DR EMBL; AB037712; BAB03520.1; -
DR EMBL; AB037713; BAB03521.1; -
DR HSP; Q45894; IELH.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR011065; Kunitz like.
DR InterPro; IPR000395; Peptidase M27.
DR InterPro; IPR006025; Pept M_Zn_BS.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOKILYSIN.
DR ProDom; PD001963; Botulinum; 1.
DR PROSITE; PS00142; ZINC PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1251 AA; 143751 MW; 2021F4E4270296 CRC64;

Query Match 34.7%; Score 2368.5; DB 2; Length 1251;
Best Local Similarity 39.94; Pred. No. 2.7e-97;
Matches 534; Conservative 253; Mismatches 418; Indels 135; Gaps 42;

Qy 1 MPFVKQFNKDPVNGVDIAVYIKIPNAGOMQPVKAFKHNKIWIPIPERDTF-TNPERGDL 59
Db 1 MPKIN-SFYNDVPNDETILYIK--PGGCEFFYKSFNIMKNIWIPIERNVIGITPQ--DF 55

Qy 60 NPPEAKQVPVSYDYSTLDNEKONYLKGVTKLPERIYSTDLGRMLLSIVRGIPFWG 119
Db 56 HPPTSLKNGDSYDYPNLYQSDBEKDFLKVITKIFNRIINNLSGGILLBELSKANLYG 115

Qy 120 G-STIDTELVIDTNCINVIQDPSYRSEELNLVIGPSADIIQFECKSFGEVLN--L 175
Db 116 NDNTPNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDL--FETNSSNISRNNYMP 173

Qy 176 TRNGYGSQVIRSPDPTFGFEESLEVDYTNPLLGAGKAFATDPVATLAHELIYAGHRLY-- 233
Db 174 SNHFGSGIALVTSPESYFRFNDN-----SMNEFIQDPALTLMHHLIHLGLYGA 224

Qy 234 -GIAPNPNRVKVN---TNAYEMSGLEVSFEELRTFGGHDAKFIQDSLONEFRLYYNK 289
Db 225 KGITTKYTIQKQNPILTN---IRGTNI--EEFLTFGGTDLNIIITSAQNDIYTNLLAD 278

Qy 290 FKDIASLTNKAISIVGTTASLQYMKNVKFEKYLISEDTSGKFSVDKLFKEDKLYKMLTEY 349
Db 279 YKTIASKLSKVQV---SNPLLNPKYQVFEAKYGLDKDASGIYSVNIKNKFNDFPKLYS-F 334

Qy 350 TEDNFVKFFKVLNKRKTYLNFDPKAVFKI-NIVPKVNTYIYDGFNLRNTNLAANGQNTIEI 408
Db 335 TEFDLATKFOVKCRQYIGQYK-YFKLSNLLNDSIYNISSGYNI--NNLKVNRFGQANL 391

Qy 409 NNNNFTKLNFTG-----LFEPYK-LLCVRGITITSKTSKSLDKGYNKALNDLCKVNW 461
Db 392 NPRIITPI---TGRGLVKKIRFECKNIIVSKIRKS-----ICIEINNGE 433

Qy 462 LFFSPSDFNTD-LANKGEI---TSDTNIEAAEENISLDLIQOYVLTFFNFENEPENIS 516
Db 462 LFFSPSDFNTD-LANKGEI---TSDTNIEAAEENISLDLIQOYVLTFFNFENEPENIS 516
```


Db 434 LFFVASENSYNDNINTPKEIDDTVTSSNNYE-----NDLQVILNFSEAP-GLS 484

Qy 517 IENLSSDIIGOLELMPNIERPNG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSV 572

Db 485 DEKLNLTQND-AVLPKYD--SNGTSDIEQHDVNLNVPFYLDQAKVPEGENNVNLTSSI 541

Qy 573 NEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWEOQLVYDFTDETSEVSTTDKIADITI 632

Db 542 DTALLEQPKIYTFSSSEFINNVNKPQVQALFVSWIQVVLVDFTTEANQKSTVDKIADISI 601

Qy 633 IIPYIGPALNIGNMLYKDDPFGALIFSGAVILLFPIEPIAIPVLGTALVSYIA---NKV 689

Db 602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLFVPELLPTILVTFIKSFLGSSDNKN 661

Qy 690 LTVQTDNALSKRNEKWDVYKYIVTNWLAKVNTQIDLRKMKKEALENAQEAATKALINY 749

Db 662 KVIRAINNALKERDEKKEVYSFVSNWMTKINTQFNKRKEQMYQALQVNAKLTIEF 721

Qy 750 QYNQYTEEEKNI--NFNIDDLSSKLNESINKAMINIKFNQCSVSYLNMSPYGVKX 807

Db 722 KYSYTLLEEKELKNYVDIEQIENELNQKVSIAWMNIDRFLTESSISYLMKLINEVKINK 781

Qy 808 LEDFDASLKDALLYVDNRGLTIGQVDR-LKDKVNNTLSTDIPOLSKYVDNQRLISTF 866

Db 782 LREYDENVKTYLLNYITQH-GSILGESQBELNSVMTDTLNNISIPFKLSSYTDKILISYF 840

Qy 867 TEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPIDQNKQIQLFNLESSKIEV 926

Db 841 NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNINNGEIFYPTNKQFTIPNSKPEVNI 900

Qy 927 ILKNAIVNSWYENFSTFWIRIPKYFNSI-SLNNEYTIINCM-ENNSGWKVSILNYGEII 984

Db 901 SQNDYIIVDNKYKNSISFWVRIPNYDNKIWNINNEYTIINCMRDNSGWKVSILNHEII 960

Qy 985 WTLQDQTEIKQVVFYKYSOMINISDYINRWLFVITNRLNNSKIYINGRLIDOKPISNL 1044

Db 961 WTLQDNARINQKLVFKYGNANGISDYINKWIFVTITNDRIGDSKLYINGHLIDOKSILNL 1020

Qy 1045 GNIHASNIMFKLDCGRDTHRYIWKYFNLPDKELNEKEIKDLYDNQSNISGILKDFWGDY 1104

Db 1021 GNIHVSNDILFKYVNCST-RYIGIRVFNIFDKELDETEIOTLYSNBPNTNLIKDFWGN 1079

Qy 1105 LOYDKPYMLMLYDPNKYVD-----VNVGIRGYMYLKGPRGSVWNTNLYNLSLYR 1156

Db 1080 LLYDKGYLLNLVLPNPNFIDRRDKDSTLSINNI-----RSTILLANRLYS 1123

Qy 1157 GTKFLIKKY-ASGNKDNIVRNDRVYINVVYKNKEYRLATNASQAGVEKILSALEIPDVG 1215

Db 1124 GIKVKIQRVNDSSTNDRFVRKNDQVYINYSNSSSYSLYADTNTTDKEKTIKS----SSSG 1180

Qy 1216 N-LSQVVMKSKDQGITNCKMQLDNNGNDIGFIGHQPNNTAKLVASWYNQRIERS 1274

Db 1181 NRPNQVVMNS-----VGNCTMFKNNNGNIGLGF----KADTVASTWYTHYTHRDH 1231

Qy 1275 SRTLGSWEFIPVDGNGER 1294

Db 1232 TNSNGCFWNPFISEHGWQEK 1251

RESULT 8

BAB86845
ID BAB86845 PRELIMINARY; PRT; 1252 AA.
AC BAB86845
DT 02-MAR-2004 (TreeBrel. 27, Created)
DT 02-MAR-2004 (TreeBrel. 27, Last sequence update)
DT 02-MAR-2004 (TreeBrel. 27, Last annotation update)
DE Botulinum neurotoxin type E.
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=35396;
RA Tsukamoto K., Mukamoto M., Kohda T., Ihara H., Wang X., Maegawa T.,
RA Nakamura S., Katasawa T., Kozaki S.;
RT "Sequence of the botulinum neurotoxin type E";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB082519; BAB86845.1; -;
KW Neurotoxin.
SQ SEQUENCE 1252 AA; 143637 MW; 76401D4D2E95D7A2 CRC64;

Query Match 34.3%; Score 2339; DB 2; Length 1252;
Best Local Similarity 39.9%; Pred. No. 5.6e-96;
Matches 535; Conservative 249; Mismatches 421; Indels 136; Gaps 43;

Qy 1 MPFVNKFNYKDPVNGVDIAYIKIPNAGQMPQKAFIHNKIWIPIRDTF-TNPEEGDL 59

Db 1 MPKIN-SFYNNDPVNDRILYIK--PGCQEFYKSFNMKNIMWIIPERNVIGTTPQ--DF 55

Qy 60 NPPEAKQVPVSYDSTYLTSDNEKONYLKVTVKLPERIYSTDLGRMLLTSIVRGIPFWG 119

Db 56 HPPTSLKNGDSYYPNQLQSDDEKDRFLKIVTKIFNRINNNLSGGILBELSKANPYLG 115

Qy 120 G-STIDTELKVIDTNCINVIQPGSYRSEELNLVIIGPSADIIQFECKSPGHEVLN---L 175

Db 116 NDNTPDNQFHIGDASAVEIKFSGSQDILLPNVIMGAEPLD--FETNSNISLRNNMP 173

Qy 176 TRNGYGSQTQIRFSPDFTFGFEESLEVDTPNLLGAGKATDPAPVLAHELIYAGHRLY-- 233

Db 174 SNHGFGSIAIVTSPESYFRFNDN-----SMNEFIQDPALTLHELHLSHLGYGA 224

Qy 234 -GTAINPNRVKVN---TNAYEMSGLEVSFEELRTFGHDAKFDLSQNEFLRYYNK 289

Db 225 KGITTKYTIQKQNPLITN---IRGTNI--EEFLTFGGTDLNIIITSAQSNDIYTNLLAD 278

Qy 290 FKDIASLANKAKSIVGTATSLQYMNQVFKKYLISEDTSGKFSVDKLFKDKLYKMLTEIY 349

Db 279 YKTIASKLSKVQV---SNPLNPKYKDVFEAKYGLDKDASGIYSVINKNFNDFPKKLYS-F 334

Qy 350 TEONFVFFKVLNKRKTYLNFDFKAVFKI-NIVPKVNYTYIDGFNLNRTNLAANFNGQTEI 408

Db 335 TEPDLATKQVKCRQTVIGQYK-YFKLSNLLNDSIYINISEGNI--NNLKVNFGQANL 391

Qy 409 NNNFTKLNKFTG-----LFEFYK-LLCVRGITTSKSLDKGYNKALNDLCIKVNNWD 461

Db 392 NPRIITPI---TGRLVKKIIRFCNKIVSVKGRKS-----ICIEINNGE 433

Qy 462 LFPSPSEDNFTND-LNKGEI-----TSDTNIIEAENISLDLQOYVLTNFDNEPENIS 516

Db 434 LFFVASENSYNDNINTPKEIDDTVTSSNNYE-----NDLQVILNFSEAP-GLS 484

Qy 517 IENLSSDIIGOLELMPNIERPNG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSV 572

Db 485 DEKLNLTQND-AVLPKYD--SNGTSDIEQHDVNLNVPFYLDQAKVPEGENNVNLTSSI 541

Qy 573 NEALLNPSRVYTFSSDYVKKVKNKATEAAMFLGWEOQLVYDFTDETSEVSTTDKIADITI 632

Db 542 DTALLEQPKIYTFSSSEFINNVNKPQVQALFVSWIQVVLVDFTTEANQKSTVDKIADISI 601

Qy 633 IIPYIGPALNIGNMLYKDDPFGALIFSGAVILLFPIEPIAIPVLGTALVSYIA---NKV 689

Db 602 VVPYIGLALNIGNEAQKGNFKDALELLGAGILLFVPELLPTILVTFIKSFLGSSDNKN 661

Qy 690 LTVQTDNALSKRNEKWDVYKYIVTNWLAKVNTQIDLRKMKKEALENAQEAATKALINY 749

Db 662 KVIRAINNALKERDEKKEVYSFVSNWMTKINTQFNKRKEQMYQALQVNAKLTIEF 721

Qy 750 QYNQYTEEEKNI--NFNIDDLSSKLNESINKAMINIKFNQCSVSYLNMSPYGVKX 807

Db 722 KYSYTLLEEKELKNYVDIEQIENELNQKVSIAWMNIDRFLTESSISYLMKLINEVKINK 781

Qy 808 LEDFDASLKDALLYVDNRGLTIGQVDR-LKDKVNNTLSTDIPOLSKYVDNQRLISTF 866

Db 782 LREYDENVKTYLLNYITQH-GSILGESQBELNSVMTDTLNNISIPFKLSSYTDKILISYF 840


```
Db 1021 GNIHVSNDILKIVNCST-YRIGIRYFNIFDKELDETEIQTLNNEBPNANILKDFWGN 1079
Qy 1105 LQYDKPYMLNLYDPNKVD-----VNVGIRGYMYLKGPRGSVMTTNIYLNSSLR 1156
Db 1080 LLYDKEYLLNVLKPNFNRTRTSLINN-----RSTILLANRLS 1123
Qy 1157 GTKFIILKY-ASGNKDNIRVNDRVYVNVVKNKEYRLATNASOAGVEKILSALIEPDVG 1215
Db 1124 GIKVKIORVNSSTNDNLVRKNDQVYINFVA-SKTHLLPLVADATTNK-EKTIKISSG 1181
Qy 1216 N-LSQVVMKSKDQGITNCKMNLQDNGNDIGFIPHQFNNAKLVASWYVNRQIERS 1274
Db 1182 NRFQVVMNS-----VGNNTMFPKNNNGNIGLGF-----KADTVVASTWYTHMRDN 1232
Qy 1275 SRTLCGSWEFIPVDGNGER 1294
Db 1233 TNSNGFFWNFISEEHGWQEK 1252

RESULT 10
Q9FAR6 PRELIMINARY; PRT; 1255 AA.
AC Q9FAR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type E botulinum toxin.
GN Name=bont/E;
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karaawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA Yamakawa K., Ogura K., Sakaguchi Y., Nakamura S.;
RT "Genetic analysis of type E botulinum toxin-producing Clostridium
RT butyricum strains.";
RL Appl. Environ. Microbiol. 66:4992-4997 (2000).
DR EMBL; AB039284; BAB1249.1; -.
DR HSP; Q45894; IE1H.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR008985; ConA_like_lect_g.
DR InterPro; IPR011065; Kunitz_like.
DR InterPro; IPR000395; Peptidase M27.
DR Pfam; PF01742; Peptidase M27.
DR PRINTS; PR00760; BONTOLIXIN.
DR PRODOM; PD001963; Botulinum; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1255 AA; 143917 MW; 1B557B9D85CD8E4D CRC64;

Query Match 34.2%; Score 2332; DB 2; Length 1255;
Best Local Similarity 40.1%; Pred. No. 1.2e-95;
Matches 537; Conservative 250; Mismatches 419; Indels 134; Gaps 43;

Qy 1 MPFVKQFNKDPVNGVDIAYIKIPNAGQMPQKAFKHKNKIWIPIPERDTF-TNPERGDL 59
Db 4 MPTIN-SFNVDPPNRTILYIK--PGCQCFYKSFNIMKNIWIPIERNVIGTIPQ--DF 58
Qy 60 NPPEAKQVPVSYDYSTLSTDNKQNYLKVTKLFRIVYSTDLGRMLLTSIVRGIPFWG 119
Db 59 LPPTSLKNGSDSYDPNYLQSDQEKDFLKVTKIFNRINDNLGRILLELSKAPYLG 118
Qy 120 G-STIDTELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIQFECKSGFHEVLN---L 175
Db 119 NDNTPDGDFFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDL--FETNSSNISLRNYP 176
```

```
Qy 176 TRNGYGSTQVIRPSPDFTQFEESLEVDTNPLLGAGKFPATDPVTLAHELIAHGRILY-- 233
Db 177 SNHFGSGIAIWTSPESYFRKDN-----SMNEFIQDPALTLAHELISHLGLYGA 227
Qy 234 -GTAINRVRPKVN--TNAYEMSGLEVSFEELRTGGHDAAKIDSLSOENEFRLYYNK 289
Db 228 KGITTKVTITOKONPLTN-----IRGTNI--EEPLTFGGTDLNIIITSAQSNDIIVTNLAD 281
Qy 290 FKDIASLTKAKSIVGTASIQVMKNVFKKYLISEDTSGKFSVDKLFKFKLYKMLTEIY 349
Db 282 YKLIASKLVQV---SNPLNPKYDVPEAKYGLDKDASGIYSVNIKNFNDIFKLYS-F 337
Qy 350 TENDVFKFKVLARKTYLNFDAVKFI-NIVPKVNYTYIDGFNLRNLTNLAANFGQNTET 408
Db 338 TEFDLATKFOVKCRQTYIGQYK-YFKLSNLLNDSIYNISEGYNI--NNLKVNFRQGANL 394
Qy 409 NNNMFTLKNFTG-----LFEPVK-LLCVRGITSTKSLDKGYNKALNDLCIKVNWMD 461
Db 395 NPRIITPI---TGRGLVKIIRFCNKIVSVKGIKRS-----ICIEINNGE 436
Qy 462 LFFSPSDNFTND-LNKGEI---TSDTNIETAAEENISLDLIQQYVLTFFNDFNEPENIS 516
Db 437 LFFVASENSYNDONINTPKEIDDTVTSNNYE-----NDLDQVILNFSEAP-GLS 487
Qy 517 IENLSSDIIQGLEMPNIERPNG-----KKYELDKYTMFHYLRAQEPFHGKSRIALTNSV 572
Db 488 DEKLNLTIQND-AVIPKYD--SNGTSDIEQHDVNLNVFFYLDQAKVPEGENNVNLTSSI 544
Qy 573 NEALLNPSRYVTFSSDYVKVKNKATEAAMPGLHVEQLVYDFDTDETSEVSTTDKIADITI 632
Db 545 DTALLEQPKIYTFSSSEFINNVKPVQALLFVGVIQQLVDFVTFTEANQKSTVDKIADISI 604
Qy 633 IIPYIGPALNIGMLYKDDFVGALIPSGAVILFIEPIAIPVLGTALYSYIA---NKV 689
Db 605 VPIYIGALNIGNEAQKGNFKDALELGCAGILLEFEPILLIPILVTIKSFLOSSDNKN 664
Qy 690 LTQVITDNLASKRNEKWEVYKIVTNWLAKVNTQIDILIRKMKKEALENQAEATKAIINY 749
Db 665 KVIAKNALKERDEKWEVYKIVSNWMTKINTQFNKRKEQMYQALQONVWALKAITES 724
Qy 750 QYNQYTEEEKNI--NPNIDDLSSKLINESINKAMININKELNQCQSVLYNSMIPYGVKR 807
Db 725 KNSYTLLEKNEKNTKYDIEQIENELNOKVSIAMNIDRELTSESSISYILKLINEVKINK 784
Qy 808 LEQFDASLKDALLKYIYDNRGTILIGQVDR-LKQKVNNTLSTDIPFQLSKYVDNORLLSTF 866
Db 785 LREYDENVKTYLLDYII-KHGSILGESQOELNSVDTLNNISPFKLSSTYDDKILISYF 843
Qy 867 TEYIKNTINTSIINLRYESNHLIDLSRYASKINIGSKVNFDPIDKQIQIQLFNLESSKIEV 926
Db 844 NKFFKRIKSSSVLNMRYKNDKYVDTSYDGNININGDVYKYPYTKNQFGIYNDKLSVNI 903
Qy 927 ILKNAIVNMYENFSTFWIRPKYFNSI-SLNNEYTIINCM-ENNSGKVSILNYGEII 984
Db 904 SONDYIYDNKYNKXNFSFVFPYIPNDYDKIVNVNNEYTIINCMRDNNNGKWSLNHNEII 963
Qy 985 WTLQDTQEIQRVVKYKYSQMINISDYINRWIFVTITNNLNNSKIYINGRLIDOKPISNL 1044
Db 964 WTLQDSGINKQKAFNAGNANGISDYINKWIFVTITNDRLDGSKLYINGNLIDKKSILNL 1023
Qy 1045 GNIHASNNIPKLDGCRDTHRYIWKYFNLFDKELNEKEIKDLYDNQSNIGILKDFWGDY 1104
Db 1024 GNIHVSNDILFKIVNCST-YRIGIRYFNIFDKELDETEIQTLNNEBPNANILKDFWGN 1082
Qy 1105 LQYDKPYMLNLYDPNKVD-----VNVGIRGYMYLKGPRGSVMTTNIYLNSSLR 1156
Db 1083 LLYDKEYLLNVLKPNFNRTRTSLINN-----RSTILLANRLS 1126
Qy 1157 GTKFIILKY-ASGNKDNIRVNDRVYVNVVKNKEYRLATNASOAGVEKILSALIEPDVG 1215
Db 1127 GIKVKIORVNSSTNDNLVRKNDQVYINFVA-SKTHLLPLVADATTNK-EKTIKISSG 1184
```


RT gene.";
 RL Eur. J. Biochem. 204:657-667(1992).
 RN [3]
 RP SEQUENCE OF 1-251 FROM N.A.
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 with other clostridial neurotoxins.";
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [4]
 RP SEQUENCE OF 1-13.
 RX MEDLINE=85197963; PubMed=3888113;
 RA Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
 RT "Partial amino acid sequences of botulinum neurotoxins types B and
 E.";
 RL Arch. Biochem. Biophys. 238:544-548(1985).
 RN [5]
 RP SEQUENCE OF 419-426.
 RX MEDLINE=90344918; PubMed=2116911;
 RA Gimenez J.A., Dasgupta B.R.;
 RT "Botulinum neurotoxin type E fragmented with endoprotease Lys-C
 reveals the site trypsin nicks and homology with tetanus neurotoxin.";
 RL Biochimie 72:213-217(1990).
 RN [6]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94063091; PubMed=8243676;
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
 Benfenati F., Wilson M.C., Montecucco C.;
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
 COOH-terminal peptide bonds.";
 RL FEBS Lett. 335:99-103(1993).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94124495; PubMed=8294407;
 RA Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
 Jahn R., Niemann H.;
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
 RL J. Biol. Chem. 269:1617-1620(1994).
 CC -|- FUNCTION: Botulinum toxin acts by inhibiting neurotransmitter
 release. It binds to peripheral neuronal synapses, is internalized
 and moves by retrograde transport up the axon into the spinal cord
 where it can move between postsynaptic and presynaptic neurons. It
 inhibits neurotransmitter release by acting as a zinc
 endopeptidase that catalyzes the hydrolysis of the 180-Arg-|-Ile-
 181 bond in SNAP-25.
 CC -|- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 detected action on small molecule substrates.
 CC -|- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -|- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
 heavy chain (H). The light chain has the pharmacological activity,
 while the N- and C-terminal of the heavy chain mediate channel
 formation and toxin binding, respectively.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- MISCELLANEOUS: There are seven antigenically distinct forms of
 botulinum neurotoxin: Types A, B, C, D, E, F, and G.
 CC -|- SIMILARITY: Belongs to peptidase family M27.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62089; CAA43999.1; -;
 DR EMBL; X62683; CAA44558.1; -;
 DR PIR; S08575; S08575.
 DR PIR; S21178; S21178.
 DR HSP; Q45894; 1EH.
 DR MEROPS; M27.002; -;
 DR InterPro; IPR008985; ConA_like_lec_gl.

DR InterPro; IPR011065; Kunitz like.
 DR InterPro; IPR000395; Peptidase_M27.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR Pfam; PF01742; Peptidase_M27; 1.
 DR PRINTS; PR00760; BONTOKILYSIN.
 DR PRODOM; PD001963; Bontokilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Direct protein sequencing; Hydrolase; Metalloprotease; Neurotoxin;
 KW Transmembrane; Zinc.
 FT INIT_MET 0
 FT CHAIN 1 421 Botulinum neurotoxin E light-chain.
 FT CHAIN 422 1250 Botulinum neurotoxin E heavy-chain.
 FT METAL 211 211 Zinc (catalytic) (By similarity).
 FT ACT_SITE 212 212 By similarity.
 FT METAL 215 215 Zinc (catalytic) (By similarity).
 FT DISULFID 411 425 Interchain (Probable).
 FT CONFLICT 176 176 R -> G (in Ref. 2).
 FT CONFLICT 197 197 C -> S (in Ref. 2 and 3).
 FT CONFLICT 339 339 R -> A (in Ref. 2).
 FT CONFLICT 772 772 I -> L (in Ref. 2).
 FT CONFLICT 962 963 FE -> LQ (in Ref. 2).
 FT CONFLICT 966 966 R -> A (in Ref. 2).
 FT CONFLICT 1194 1194 N -> NN (in Ref. 2).
 SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;
 Query Match 33.9%; Score 2309.5; DB 1; Length 1250;
 Best Local Similarity 39.6%; Pred. No. 1-2e-94;
 Matches 530; Conservative 250; Mismatches 423; Indels 137; Gaps 43;
 QY 2 PFVKNQPNYKDPVNGVDIAIYKIPNAGQOMQVPKAFKHNNKIWIPIERDTP-TNPEEGDLN 60
 DB 1 PKIN-SFNYNDPVNDRTILYIK-PGCGCFYKSFNIMKNIWIPIERNVIGTTPQ--DFH 55
 QY 61 PPEAKQVPVSYVDSTYSLTDNEKDNVYKVTXLPRIYSTDIGRMILTISVIRGIPWGG 120
 DB 56 PPTSLKNGDSSYYDPNYLOSDEEKDRFLKTVTKFNRRNNLGGILLLELSKANPYLGN 115
 QY 121 -STIDYELKVIDTNCINVIQDGSYRSEELNLVIIGPSADIIQFECKSFGEVLN--LT 176
 DB 116 DNTPDQNFHIGDASAVEIKFESNGSQDILLPNVILMGAEPDL--FETNSSNISLRNNWPS 173
 QY 177 RNYGSGTQYIRSPDPFTFGPEESLEVDVNTNPLLGAGKATDPATVLAHELIIYAGHRLY--- 233
 DB 174 NHRFGSIAITVTFSPESYFRFENDNC-----MNEFIQDPALTLMHELIHSLHLYGAK 224
 QY 234 GIAINPNRVFKVN--TNAYENSGLVSPFEELRTFGGHDAKFIDSIQENEFRLYYNKF 290
 DB 225 GITTKYITQKQPLITN-----IRGTNI--EELTFGGTDLNITSAQSNDIITNLADY 278
 QY 291 KDIASLTINKAKSIVGTTSASIQYMKNVFKKYLSEDTSGKFSVDKLFKDKLYKMLTEIYT 350
 DB 279 KTIASKLSKVQV--SNPLNPKYDVEAKYGLDKDASGIYSVNINKNFDFKKLYS-FT 334
 QY 351 EDNPFVFFKVLNKRKTYLNFDKAVFKI-NIVPKVNYTYDGFNLRNNTNLAANFNGQNTIN 409
 DB 335 EFDLRTKFQVKCRQTYIGQYK-YFKLSNLINDSIYINISEGYNI--NNLKVNFRGQANLN 391
 QY 410 NMNFTKLKNFTG-----LPEFYK-LLCYVRGIITSTKSLDKGNKALNDLCIKVNNWDL 462
 DB 392 PRIITPI--TGRLVKKIRFCNKIVSVKGIKRS-----ICIEINNGEL 433
 QY 463 FFSFSDNFTND-LNKGEET---TSDTNI EAENISLDLIQQYYLTTFNFDNPEPENISI 517
 DB 434 FFVASENSYNDNDINTPKETDDTDTVSNNVE-----NDLDQVILNFNSESAP-GLSD 484
 QY 518 ENLSSDIIGLELMPNIERFPNG-----KKYELDKYTMFHYLRAQEFHGKSRALTNSVN 573
 DB 485 EKLNLATQND-AIIPKYD--SNGTSDIEQHDVNELVNFFYLDQAQKVPGESENNYLTSSID 541
 QY 574 EALINPFRVYTFSSDYVKKVNKATEAAMFLGWVEQLVYDFDTDETSEVSTTDKIADITII 633
 DB 542 TALLEQPKIYTFSSSEFINNVNKPQALFVSWIQQVLVDFTTEANQKSTVDKIADISIV 601

Db	927	NSRYQNFISGFVTRIPKHYIPDMHNREYTTINCMNNNSGWSKISLRVTVRDCEIITWLTQDT	988
Qy	991	QELKQRVFYKISQMINISDYINRWIFVTYITNRLNNSKIYINGELIDQKPSINLGNTHAS	1050
Db	987	SGNKENLIPRYEELNRIISNINKWIFVTYITNRLNLSRIYINGNLIVEKISINLGDHVS	1046
Qy	1051	NNIMFKLDCGDTHRYIWKYFNLFDKELAKEIKDLVDNQSNGILKDFWGDYLOYDKP	1110
Db	1047	DNILFKIVGC-DDETYYGIRYFKVFNTLDEKTEIETLVSNEPDSILKNYWGNYLLYNKK	1105
Qy	1111	YYMLNLDPNKKYVDVNVNNGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFTIKKYAS---	1167
Db	1106	YYLFNLRLKDKYITLNS-----GILNINQORG-VTEGSVFLNYKLYEGVEVIIRKNGPDI	1160
Qy	1168	GKNKDINVRNDRVYINVVKNKCYRLATNASQAGVEKILSALETDPDGNLSQVVMKSKN	1227
Db	1161	SNTDNFVRKNDLAVINVDRGEVRYLAD-TKSEKIKIIRTSNLND--SLGQIIIVMS--	1215
Qy	1228	DOGITKCKKNLODNGNDIGFTGFHCFNNIAKIVASWYNNRQIERSRSLTGCSSWEFTPV	1287
Db	1216	---IGNCTWVFQNNNGSNLGLGFHNN-----LVASSWYNNIRRTSSNGCFWSSLSK	1268
Qy	1288	DDGWGE 1293	
Db	1269	ENGWKE 1274	
RESULT 15			
Qy	Q8GR96	PRELIMINARY; PRT; 1291 AA.	
AC	Q8GR96;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Neurotoxin.		
GN	Name:bontx.		
OS	Clostridium botulinum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCHI_TaxID=1491;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Inara H., Kohda T., Morimoto P., Tsukamoto K., Karasawa T.,		
RA	Nakamura S., Mukamoto M., Kozaki S.;		
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.		
EMBL	AB084152; BAC22064.1; -		
DR	HSP; P10844; IEPW.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0009405; P:pathogenesis; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR011591; Botulinum.		
DR	InterPro; IPR008985; ConA like lec_gl.		
DR	InterPro; IPR011065; Kunitz like.		
DR	InterPro; IPR000395; Peptidase M27.		
DR	InterPro; IPR006025; Pept_M_Zn_BS.		
DR	Pfam; PF01742; Peptidase_M27; 1.		
DR	PRINTS; PR00760; BONTOKILYSIN.		
DR	ProDom; PD001963; Botulinum; 1.		
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.		
KW	Neurotoxin.		
SEQ	SEQUENCE 1291 AA; 150574 MW; 0227CAEF4F58504D CRC64;		
Query Match 33.6%; Score 2290; DB 2; Length 1291;			
Best Local Similarity 39.1%; Pred.No.8.9e-94;			
Matches 523; Conservative 232; Mismatches 490; Indels 94; Gaps 25;			
Qy	1	MPFVNKFQNVKDPVNGVDIAVIKIPNA-GQMPVKAFKHINKIWIIPERTFF-TNPBGD	58
Db	1	MPVTNNFNVNDIDNNNIIMWEPFPARGTGRYYKAFKIDRIWIIPERYTGYKPEDFN	60
Qy	59	LNPPEAKQVPVSYDYSTYLSTDNEKNYKLVGVTKLPERIVSTDLGRMLLTSGVIRGIPFW	118
Db	61	KSSGIFNRDY-CBYDYPDLNTNDKKNIQFTWIKLPIKSKPLGKLEMLINGIPYL	119

This Page Blank (uspto)